GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 6, 2004, 08:32:42 ; Search time 52 Seconds
(without alignments)
48.902 Million cell updates/sec

Title: US-09-458-302B-193
Perfect score: 41
Sequence: 1 IMIGVLVGV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	9	2	AAW70048	Aaw70048 CEA deriv
2	41	100.0	9	2	AAY47657	Aay47657 Immunogen
3	41	100.0	9	4	AAB99695	Aab99695 HLA A2 bi
4	41	100.0	9	4	AAG62397	Aag62397 Immunogen
5	41	100.0	9	5	AAU95888	Aau95888 Immunogen
6	41	100.0	9	6	AAE35579	Aae35579 Human CAE
7	41	100.0	9	6	ABU04842	Abu04842 Human exp
8	41	100.0	9	6	ABU04846	Abu04846 Human exp
9	41	100.0	9	6	ABU03347	Abu03347 Human exp
10	41	100.0	9	6	ABU04844	Abu04844 Human exp
11	41	100.0	9	6	ABG74919	Abg74919 melanoma-
12	41	100.0	9	7	ADD84718	Add84718 Human car
13	41	100.0	10	2	AAW70047	Aaw70047 CEA deriv
14	41	100.0	10	2	AAW70051	Aaw70051 CEA deriv
15	41	100.0	10	2	AAY47672	Aay47672 Immunogen
16	41	100.0	10	2	AAY47668	Aay47668 Immunogen
17	41	100.0	10	5	AAU95892	Aau95892 Immunogen
18	41	100.0	10	5	AAU95889	Aau95889 Immunogen
19	41	100.0	10	6	ABU04847	Abu04847 Human exp
20	41	100.0	10	6	ABU04845	Abu04845 Human exp
21	41	100.0	10	6	ABU04848	Abu04848 Human exp
22	41	100.0	10	6	ABU04843	Abu04843 Human exp
23	41	100.0	12	6	AAE35583	Aae35583 Human CAE
24	41	100.0	12	6	AAE35585	Aae35585 Human CAE
25	41	100.0	12	6	AAE35584	Aae35584 Human CAE
26	41	100.0	13	6	AAE35582	Aae35582 Trypsin d
27	41	100.0	16	7	ADD84716	Add84716 Human car
28	41	100.0	21	6	AAE35580	Aae35580 TA compri
29	41	100.0	21	6	AAE35586	Aae35586 TA compri
30	41	100.0	29	2	ABU10423	Abu10423 Urokinase
31	41	100.0	227	5	AAE24334	Aae24334 Human lun
32	41	100.0	372	6	ABU04802	Abu04802 Human exp
33	41	100.0	493	2	AAR77435	Aar77435 BGP (1-31
34	41	100.0	595	4	AAG74097	Aag74097 Human col
35	41	100.0	595	6	ABU04813	Abu04813 Human exp
36	41	100.0	670	5	AAU98922	Aau98922 Human car
37	41	100.0	697	6	ABU04809	Abu04809 Human exp
38	41	100.0	698	1	AAP81229	Aap81229 Carcinoem
39	41	100.0	698	2	AAR65168	Aar65168 Carcinoem
40	41	100.0	698	2	AAW22844	Aaw22844 Human car
41	41	100.0	698	6	ABU04822	Abu04822 Human exp
42	41	100.0	698	6	ABU04806	Abu04806 Human exp
43	41	100.0	698	6	ABU04816	Abu04816 Human exp
44	41	100.0	698	6	ABU04798	Abu04798 Human exp
45	41	100.0	698	6	ABU04827	Abu04827 Human exp

ALIGNMENTS

RESULT 1
AAW70048
ID AAW70048 standard; peptide; 9 AA.
XX
AC AAW70048;
XX
DT 22-OCT-1998 (first entry)
XX
DE CEA derived HLA-A2.1 binding peptide 5 (residues 691-699).
XX
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9833888-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US001959.
XX
PR 31-JAN-1997; 97US-0036696P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Tsai V, Southwood S, Sidney J, Sette A, Celis E;
XX
DR WPI; 1998-437445/37.
XX
PT Production of antigen-specific cytotoxic T cells - by incubating
PT immunogenic peptide(s) from antigen that binds class I major
PT histocompatibility complex molecules with pre-treated antigen presenting
PT cells.
XX
PS Example 6; Page 75; 104pp; English.
XX
CC Sequences shown in AAW70044 to AAW70052 represent peptides derived from
CC carcinoembryonic antigen (CEA). The peptides can bind to a human
CC leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC vitro. The method comprises contacting immunogenic peptides from an
CC antigen that binds class I major histocompatibility complex (MHC)
CC molecules with antigen presenting cells (APCs) pretreated with
CC pretreatment growth factors, and incubating the APCs with purified CD8

CC cells in the presence of at least 2 incubation growth factors, thereby
CC producing antigen-specific CTLs. A method for specifically killing target
CC cells in a human patient is also provided which comprises obtaining a
CC fluid sample containing CTLs from a patient, contacting the cytotoxic T
CC cells with APCs pretreated with pre-treatment growth factors, where the
CC APCs comprise class I MHC molecules. The pretreated APCs are incubated
CC with the cytotoxic growth factors, thereby producing activated CTLs which
CC are contacted with a carrier to form a composition. The composition can
CC then be administered to the patient. The activated CTLs can be used for
CC treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 41; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|||||||||
Db 1 IMIGVLVGV 9

RESULT 2
AAY47657
ID AAY47657 standard; peptide; 9 AA.
XX
AC AAY47657;
XX
DT 01-DEC-1999 (first entry)
XX
DE Immunogenic peptide having a human leukocyte antigen binding motif #2268.
XX
KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW immune response; T cell activation; major histocompatibility complex;
KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW vaccine; immunisation.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9945954-A1.
XX
PD 16-SEP-1999.
XX
PF 13-MAR-1998; 98WO-US005039.
XX
PR 13-MAR-1998; 98WO-US005039.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Kubo RT, Sidney J, Celis E, Grey HM, Southwood S;
XX
DR WPI; 1999-551214/46.
XX
PT New immunogenic peptides with HLA binding motif, useful in treatment and
PT diagnosis of cancers and viral diseases.
XX
PS Claim 1; Page 118; 150pp; English.
XX
CC AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
CC having a human major histocompatibility complex (MHC) Class I (also known
CC as human leukocyte antigen (HLA)) binding motif. The immunogenic peptides
CC can bind to a specific HLA allele (i.e. HLA-A subtypes HLA-A2.1, A1, A3.2
CC or A24.1 or HLA-B or C) and induce a cytotoxic T cell response against
CC the antigen from which the peptide is derived. Cytotoxic T lymphocytes
CC (CTLs) which destroy antigen-bearing cells are normally induced by an
CC antigen in the form of a peptide fragment bound to a HLA molecule, rather
CC than the intact foreign antigen itself, and are particularly important in
CC tumour rejection and in fighting viral infections. The peptides are
CC therefore useful therapeutically to treat or prevent viral infections and
CC cancers in mammals (especially humans) e.g. prostate cancer, hepatitis B
CC and C, AIDS, and renal carcinoma. They can be administered as vaccines to
CC elicit an immune response in individuals susceptible or otherwise at risk
CC of viral infection or cancer, or used to treat chronic or acute
CC conditions. They are also useful diagnostically, and can be used to
CC induce a cytotoxic T cell response, by contacting a cytotoxic T cell with
CC the peptide e.g. to produce CTLs ex vivo for infusion back into a
CC patient. The polynucleotides encoding the immunogenic peptides are also
CC useful therapeutically and for immunisation as above
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 41; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|||||||||
Db 1 IMIGVLVGV 9

RESULT 3
AAB99695
ID AAB99695 standard; peptide; 9 AA.
XX
AC AAB99695;
XX
DT 06-SEP-2001 (first entry)
XX
DE HLA A2 binding CTL epitope peptide from CEA SEQ ID NO:16.
XX
KW Human leukocyte antigen A2 binding peptide; HLA class I A2; CTL;
KW cytotoxic T-cell lymphocyte; tumour associated antigen; CEA; HER2/neu;
KW MAGE2; MAGE3; p53; vaccine; cancer; cytostatic; immunomodulator;
KW immunotherapy; immune response.
XX
OS Homo sapiens.
XX
PN WO200141741-A1.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-US034318.
XX
PR 13-DEC-1999; 99US-0170448P.
PR 05-APR-2000; 2000US-00543608.
PR 30-MAY-2000; 2000US-00583200.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Celis E, Keogh E;
PI Chesnut R;
XX
DR WPI; 2001-381489/40.
XX
PT Compositions for use in a vaccine for treating, e.g., breast, lung and
PT colon cancer comprises at least one peptide that comprises an isolated
PT epitope of a tumor-associated antigen.
XX
PS Claim 1; Page 76; 86pp; English.
XX
CC The present invention describes a composition (I) comprising at least one
CC peptide that comprises an isolated, prepared epitope consisting of a
CC sequence selected from 25 short amino acid sequences given in AAB99680 to
CC AAB99704. Also described are: (1) a composition (II) comprising one or
CC more peptides, and further comprising at least two epitopes selected from
CC the 25 short amino acid sequences (as above), where each of the one or
CC more peptides comprise less than 50 contiguous amino acids that have 100%
CC identity with a native peptide sequence; and (2) a vaccine composition
CC (III) comprising an epitope selected from the 25 short amino acid
CC sequences (as above) and a pharmaceutical excipient. (I) has cytostatic
CC and immunomodulatory activities and can be used in vaccine production and
CC immunotherapy. The peptide epitope compositions (I)-(II) are useful for
CC monitoring an immune response to a tumour associated antigen or when one

```
CC   or more peptides are combined to create a vaccine (III) that stimulates
CC   the cellular arm of the immune system. In particular, the vaccine
CC   mediates immune responses against tumours in individuals who bear an
CC   allele of the human leukocyte antigen (HLA)-A2 supertype and improve the
CC   standard of care for patients being treated for breast, colon, or lung
CC   cancer
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 41;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 4
AAG62397
ID   AAG62397 standard; peptide; 9 AA.
XX
AC   AAG62397;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Immunogenic peptide CEA.691 SEQ ID 1.
XX
KW   Class I epitope; immunogenic; heteroclitic analogue; immune response;
KW   antigen display; viral disease; cancer.
XX
OS   Synthetic.
XX
PN   WO200136452-A2.
XX
PD   25-MAY-2001.
XX
PF   20-NOV-2000; 2000WO-US031856.
XX
PR   18-NOV-1999;   99US-0166529P.
PR   06-OCT-2000; 2000US-0239008P.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Tangri S,  Sette A,  Ishioka G;
XX
DR   WPI; 2001-355609/37.
XX
PT   Enhancing immunogenicity of peptide containing class I epitope, useful
PT   for treating cancer, comprises providing (semi-)conservative amino acid
PT   substitutions at specified positions of these epitopes.
XX
PS   Example 1; Fig 1A; 96pp; English.
XX
CC   This invention relates to a method of enhancing the immunogenicity of a
CC   peptide, where the peptide contains a class I epitope. The invention
CC   includes methods for preparing peptides containing epitopes which have
CC   enhanced ability to effect an immune response (compared to wild-type
CC   epitopes). The peptides are referred to as heteroclitic analogues. The
CC   method is useful for eliciting an immune response by contacting CTLs with
CC   the immunogenically enhanced peptide in vitro in the presence of an
CC   antigen presenting cell, or by administering to a subject a nucleic acid
CC   molecule comprising a nucleotide sequence encoding the peptide. The
CC   peptides are useful as reagents to evaluate an immune response and the
CC   efficacy of the vaccine, and for making antibodies. The heteroclitic
CC   analogues are useful in immunological compositions for the treatment of
CC   viral diseases, cancer, and other conditions which are characterised by
CC   displayed antigens on target cells. The present sequence represents a
CC   class I epitope which may be used in the method of the invention
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 41;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 5
AAU95888
ID   AAU95888 standard; peptide; 9 AA.
XX
AC   AAU95888;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Immunogenic peptide with (HLA)-A2.1 binding site #101.
XX
KW   HLA-A2.1 binding peptide; cytostatic; virucide; anti-HIV; hepatotropic;
KW   human immunodeficiency virus; antiinflammatory; antibacterial; vaccine;
KW   protozoacide; immunosuppressant; immunogenic peptide; T cell activation;
KW   human leucocyte antigen binding site; cytotoxic T cell response;
KW   viral infection; hepatitis; Epstein-Barr virus; papilloma virus;
KW   human immunodeficiency virus; HIV; Kaposi sarcoma; Lassa fever virus;
KW   cytomegalovirus; tumour; prostate cancer; renal carcinoma; lymphona;
KW   prostate-specific antigen; p53; carcino-embryonal antigen;
KW   melanoma antigen; Mycobacterium tuberculosis; protozoa;
KW   trypanosome surface antigen; condlyloma acuminatum.
XX
OS   Unidentified.
XX
PN   WO200220616-A1.
XX
PD   14-MAR-2002.
XX
PF   01-SEP-2000; 2000WO-US024102.
XX
PR   01-SEP-2000; 2000WO-US024102.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Grey HM,  Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 2002-351766/38.
XX
PT   Immunogenic peptide with human leucocyte antigen-A2.1 binding site,
PT   useful for treating e.g. viral infection or tumors.
XX
PS   Claim 1; Page 27; 35pp; English.
XX
CC   The invention describes a composition comprising an immunogenic peptide
CC   having a human leucocyte antigen (HLA)-A2.1 binding site. The peptides
CC   bind specifically to HLA-A2.1, to cause T cell activation and thus a
CC   cytotoxic T cell response. The peptides and the nucleic acids that
CC   encodes them, are used, in vivo or ex vivo, for treatment of viral
CC   infections (hepatitis B or C; Epstein-Barr; human immune deficiency;
CC   Kaposi sarcoma; human papilloma; Lassa fever or cytomegaloviruses);
CC   tumours including prostate cancer, renal carcinoma and lymphona (where
CC   directed to prostate-specific antigen, p53, carcino-embryonal antigen,
CC   Her2/neu or melanoma antigens); infection by Mycobacterium tuberculosis
CC   or protozoa (directed to trypanosome surface antigen); and condlyloma
CC   acuminatum. The peptides are suitable for use in peptide-based vaccines.
CC   This sequence represents an immunogenic peptide with the human leucocyte
CC   antigen (HLA)-A2.1 binding site, described in the invention
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 41;  DB 5;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
```

```
Db          1 IMIGVLVGV 9


RESULT 6
AAE35579
ID   AAE35579 standard; peptide; 9 AA.
XX
AC   AAE35579;
XX
DT   17-JUN-2003  (first entry)
XX
DE   Human CAE specific HLA-A2-restricted minimal CTL epitope #1.
XX
KW   Fusion agent; immunogenic; proliferative disease; infectious disease;
KW   cancer; therapy; vaccine; melanoma; carcinoembryonic antigen; CAE; TA;
KW   Trojan antigen; human; HLA-A2-restricted CTL epitope.
XX
OS   Homo sapiens.
XX
PN   WO200294994-A2.
XX
PD   28-NOV-2002.
XX
PF   20-MAY-2002; 2002WO-US015992.
XX
PR   18-MAY-2001; 2001US-0291874P.
XX
PA   (MAYO-) MAYO FOUND MEDICAL EDUCATION RES.
XX
PI   Celis E;
XX
DR   WPI; 2003-140367/13.
XX
PT   Fusion agent useful for preventing and treating an infectious disease, or
PT   a proliferative disease, such as cancer, comprises a transport domain,
PT   two cleavage sites, a peptide epitope and a biologically active agent.
XX
PS   Example 2; Page 43; 72pp; English.
XX
CC   The invention relates to a fusion agent (Trojan antigen; TA) comprising a
CC   transport domain, two cleavage sites, a peptide epitope recognised by an
CC   antigen-specific receptor on an effector T-lymphocyte precursor cell and
CC   a biologically active agent, where there is a cleavage site between the
CC   peptide epitope and the biologically active agent and between each
CC   biologically active agent. The fusion agent is used to make a cell
CC   immunogenic or antigenic. It is also useful for preventing and treating
CC   an infectious disease such as viral, bacterial, protozoal, fungal or
CC   yeast disease, or proliferative disease such as cancer (e.g. melanoma,
CC   neural tissue, gastrointestinal, breast, lung, ovarian, testicular,
CC   prostate, cervical, bladder, vaginal, liver, renal, bone, haematological
CC   or vascular tissue cancer). The invention is used as vaccines. The
CC   present sequence is human carcinoembryonic antigen (CAE) specific HLA-A2-
CC   restricted minimal CTL epitope. This peptide is used in the
CC   exemplification of the invention
XX
SQ   Sequence 9 AA;

  Query Match           100.0%;  Score 41;  DB 6;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 7
ABU04842
ID   ABU04842 standard; protein; 9 AA.
XX
AC   ABU04842;
XX
DT   29-JAN-2003  (first entry)
XX
DE   Human expressed protein tag (EPT) #1508.
XX
KW   Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW   protease; protease inhibitor; transporter; cytoskeletal protein;
KW   receptor; transcription factor; cancer; MHC;
KW   major histocompatability complex; myeloma; colon cancer; gastric cancer;
KW   adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200278524-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US009671.
XX
PR   28-MAR-2001; 2001US-0279495P.
PR   21-MAY-2001; 2001US-0292544P.
PR   08-AUG-2001; 2001US-0310801P.
PR   01-OCT-2001; 2001US-0326370P.
PR   04-DEC-2001; 2001US-0336780P.
PR   20-FEB-2002; 2002US-0358985P.
XX
PA   (ZYCO-) ZYCOS INC.
XX
PI   Chicz RM,  Tomlinson AJ,  Urban RG;
XX
DR   WPI; 2003-040607/03.
XX
PT   New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT   cytoskeletal proteins, receptors or transcription factors), useful for
PT   treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT   leukemia.
XX
PS   Example 2; SEQ ID NO 1508; 134pp; English.
XX
CC   The invention describes a purified polypeptide, which comprises a
CC   fragment of a kinase, phosphatase, protease, protease inhibitor,
CC   transporter, cytoskeletal protein, receptor or transcription factor. The
CC   polypeptide is useful as an immunogenic composition for eliciting in a
CC   mammal an immunogenic response directed against any of the purified
CC   polypeptide. The purified polypeptide, or the antibody that binds to this
CC   polypeptide, is useful for treating cancer. The polypeptide is also
CC   useful for identifying compounds that binds to a naturally processed
CC   class I or class II MHC-binding polypeptide. The polypeptides and
CC   polynucleotides are particularly useful for treating or preventing
CC   myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC   lymphoma or leukaemia. These are also useful for screening agents for
CC   treating the above mentioned diseases. This sequence represents an
CC   expressed protein tag (EPT) isolated from human tissue for translational
CC   profiling. Note: This sequence does not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 9 AA;

  Query Match           100.0%;  Score 41;  DB 6;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 8
ABU04846
ID   ABU04846 standard; protein; 9 AA.
XX
AC   ABU04846;
XX
DT   29-JAN-2003  (first entry)
```

```
XX
DE   Human expressed protein tag (EPT) #1512.
XX
KW   Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW   protease; protease inhibitor; transporter; cytoskeletal protein;
KW   receptor; transcription factor; cancer; MHC;
KW   major histocompatability complex; myeloma; colon cancer; gastric cancer;
KW   adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200278524-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US009671.
XX
PR   28-MAR-2001; 2001US-0279495P.
PR   21-MAY-2001; 2001US-0292544P.
PR   08-AUG-2001; 2001US-0310801P.
PR   01-OCT-2001; 2001US-0326370P.
PR   04-DEC-2001; 2001US-0336780P.
PR   20-FEB-2002; 2002US-0358985P.
XX
PA   (ZYCO-) ZYCOS INC.
XX
PI   Chicz RM,  Tomlinson AJ,  Urban RG;
XX
DR   WPI; 2003-040607/03.
XX
PT   New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT   cytoskeletal proteins, receptors or transcription factors), useful for
PT   treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT   leukemia.
XX
PS   Example 2; SEQ ID NO 1512; 134pp; English.
XX
CC   The invention describes a purified polypeptide, which comprises a
CC   fragment of a kinase, phosphatase, protease, protease inhibitor,
CC   transporter, cytoskeletal protein, receptor or transcription factor. The
CC   polypeptide is useful as an immunogenic composition for eliciting in a
CC   mammal an immunogenic response directed against any of the purified
CC   polypeptide. The purified polypeptide, or the antibody that binds to this
CC   polypeptide, is useful for treating cancer. The polypeptide is also
CC   useful for identifying compounds that binds to a naturally processed
CC   class I or class II MHC-binding polypeptide. The polypeptides and
CC   polynucleotides are particularly useful for treating or preventing
CC   myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC   lymphoma or leukaemia. These are also useful for screening agents for
CC   treating the above mentioned diseases. This sequence represents an
CC   expressed protein tag (EPT) isolated from human tissue for translational
CC   profiling. Note: This sequence does not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 9 AA;

  Query Match           100.0%;  Score 41;  DB 6;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 9
ABU03347
ID   ABU03347 standard; protein; 9 AA.
XX
AC   ABU03347;
XX
DT   29-JAN-2003  (first entry)
```

```
XX
DE   Human expressed protein tag (EPT) #127.
XX
KW   Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW   protease; protease inhibitor; transporter; cytoskeletal protein;
KW   receptor; transcription factor; cancer; MHC;
KW   major histocompatability complex; myeloma; colon cancer; gastric cancer;
KW   adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200278524-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US009671.
XX
PR   28-MAR-2001; 2001US-0279495P.
PR   21-MAY-2001; 2001US-0292544P.
PR   08-AUG-2001; 2001US-0310801P.
PR   01-OCT-2001; 2001US-0326370P.
PR   04-DEC-2001; 2001US-0336780P.
PR   20-FEB-2002; 2002US-0358985P.
XX
PA   (ZYCO-) ZYCOS INC.
XX
PI   Chicz RM,  Tomlinson AJ,  Urban RG;
XX
DR   WPI; 2003-040607/03.
XX
PT   New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT   cytoskeletal proteins, receptors or transcription factors), useful for
PT   treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT   leukemia.
XX
PS   Claim 10; SEQ ID NO 127; 134pp; English.
XX
CC   The invention describes a purified polypeptide, which comprises a
CC   fragment of a kinase, phosphatase, protease, protease inhibitor,
CC   transporter, cytoskeletal protein, receptor or transcription factor. The
CC   polypeptide is useful as an immunogenic composition for eliciting in a
CC   mammal an immunogenic response directed against any of the purified
CC   polypeptide. The purified polypeptide, or the antibody that binds to this
CC   polypeptide, is useful for treating cancer. The polypeptide is also
CC   useful for identifying compounds that binds to a naturally processed
CC   class I or class II MHC-binding polypeptide. The polypeptides and
CC   polynucleotides are particularly useful for treating or preventing
CC   myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC   lymphoma or leukaemia. These are also useful for screening agents for
CC   treating the above mentioned diseases. This sequence represents an
CC   expressed protein tag (EPT) isolated from human tissue for translational
CC   profiling. Note: This sequence does not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 9 AA;

  Query Match           100.0%;  Score 41;  DB 6;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 10
ABU04844
ID   ABU04844 standard; protein; 9 AA.
XX
AC   ABU04844;
XX
DT   29-JAN-2003  (first entry)
```

```
XX
DE   Human expressed protein tag (EPT) #1510.
XX
KW   Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW   protease; protease inhibitor; transporter; cytoskeletal protein;
KW   receptor; transcription factor; cancer; MHC;
KW   major histocompatability complex; myeloma; colon cancer; gastric cancer;
KW   adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200278524-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US009671.
XX
PR   28-MAR-2001; 2001US-0279495P.
PR   21-MAY-2001; 2001US-0292544P.
PR   08-AUG-2001; 2001US-0310801P.
PR   01-OCT-2001; 2001US-0326370P.
PR   04-DEC-2001; 2001US-0336780P.
PR   20-FEB-2002; 2002US-0358985P.
XX
PA   (ZYCO-) ZYCOS INC.
XX
PI   Chicz RM,  Tomlinson AJ,  Urban RG;
XX
DR   WPI; 2003-040607/03.
XX
PT   New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT   cytoskeletal proteins, receptors or transcription factors), useful for
PT   treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT   leukemia.
XX
PS   Example 2; SEQ ID NO 1510; 134pp; English.
XX
CC   The invention describes a purified polypeptide, which comprises a
CC   fragment of a kinase, phosphatase, protease, protease inhibitor,
CC   transporter, cytoskeletal protein, receptor or transcription factor. The
CC   polypeptide is useful as an immunogenic composition for eliciting in a
CC   mammal an immunogenic response directed against any of the purified
CC   polypeptide. The purified polypeptide, or the antibody that binds to this
CC   polypeptide, is useful for treating cancer. The polypeptide is also
CC   useful for identifying compounds that binds to a naturally processed
CC   class I or class II MHC-binding polypeptide. The polypeptides and
CC   polynucleotides are particularly useful for treating or preventing
CC   myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC   lymphoma or leukaemia. These are also useful for screening agents for
CC   treating the above mentioned diseases. This sequence represents an
CC   expressed protein tag (EPT) isolated from human tissue for translational
CC   profiling. Note: This sequence does not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 41;  DB 6;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 11
ABG74919
ID   ABG74919 standard; peptide; 9 AA.
XX
AC   ABG74919;
XX
DT   11-JUL-2003  (first entry)
XX
DE   melanoma-associated antigen MART-1 associated peptide.
XX
KW   Dendritic cell; cell line; CD124; CD116; cytostatic; antirheumatic;
KW   immunosuppressive; immunostimulatory; antibacterial; virucide;
KW   antiparasitic; fungicide; dermatological; antiinflammatory; antianaemic;
KW   nephrotropic; thyrotropic; antidiabetic; anthelmintic;
KW   protozoacide allogenic; immunotherapeutic; humoral immune system;
KW   cellular immune system; natural killer cell; CD4+ cell;
KW   antigen cytotoxic T cell; proliferation; vaccine; infection; tumour;
KW   autoimmune disease; Hashimoto's syndrome; insulin-dependent diabetes;
KW   rheumatism; systemic lupus erythematosus; Goodpasture syndrome;
KW   transplantation; melanoma-associated antigen; MART-1.
XX
OS   Homo sapiens.
XX
PN   WO2003023023-A1.
XX
PD   20-MAR-2003.
XX
PF   19-AUG-2002; 2002WO-EP009260.
XX
PR   17-AUG-2001; 2001DE-01039428.
XX
PA   (NEMO-) NEMOD IMMUNTHERAPIE AG.
XX
PI   Goletz S,  Scheper RJ,  Masterson A,  Pinedo HM;
XX
DR   WPI; 2003-301068/29.
XX
PT   Preparation of dendritic cells, useful e.g. as antitumor or antimicrobial
PT   vaccines, by treating CD124- and CD116-positive cells with stimulatory
PT   molecules.
XX
PS   Disclosure; Page 35; 89pp; German.
XX
CC   This invention describes a novel method for preparing effective dendritic
CC   cells or cell lines comprising treating cells of CD124- and CD116-
CC   positive lines with at least one stimulatory molecule, applied at the
CC   same time or sequentially. The products of the invention have cytostatic,
CC   antirheumatic,immunosuppressive, immunostimulatory, antibacterial,
CC   virucide, antiparasitic, fungicide, dermatological, antiinflammatory,
CC   antianaemic, nephrotropic, thyrotropic, antidiabetic, anthelmintic and
CC   protozoacide activity. The novel cell lines are useful: (a) as
CC   (semi)allogenic immunotherapeutic agents; (b) for activating, inhibiting
CC   or modulating the humoral and/or cellular immune systems; (c) for
CC   stimulating natural killer, CD4+ and/or cytotoxic T cells; (d) for
CC   processing and presenting antigens; and (e) to induce proliferation of
CC   immune cells. Particularly they are used to treat or prevent, as
CC   vaccines, infections (by viruses, bacteria, parasites, protozoa, prions
CC   or helminths), tumours and/or autoimmune diseases (e.g. anaemia;
CC   Hashimoto's syndrome; insulin-dependent diabetes; rheumatism; systemic
CC   lupus erythematosus; Goodpasture syndrome and many others listed); also
CC   in transplantation medicine and for diagnosis. Systems containing novel
CC   cell lines are also useful for testing the immuno-activating, -inhibiting
CC   and/or -modulating activities of substances and/or for analyzing the
CC   biology of dendritic cells. This sequence represents a peptide associated
CC   with the melanoma-associated antigen MART-1 which is described in the
CC   disclosure of the invention
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 41;  DB 6;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db          1 IMIGVLVGV 9


RESULT 12
ADD84718
```

ID ADD84718 standard; peptide; 9 AA.
XX
AC ADD84718;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human carcinoembryonic antigen (CEA) epitope peptide SEQ ID NO:7.
XX
KW identification;
KW class I major histocompatibility complex-binding fragment;
KW class I MHC molecule; class I MHC-binding fragment; cytostatic; cancer;
KW human; carcinoembryonic antigen; CEA; epitope.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO2003082317-A1.
XX
PD 09-OCT-2003.
XX
PF 20-MAR-2003; 2003WO-US008427.
XX
PR 22-MAR-2002; 2002US-0366822P.
XX
PA (ZYCO-) ZYCOS INC.
PA (AVET) AVENTIS PASTEUR INC.
XX
PI Chicz RM, Tomlinson AJ;
XX
DR WPI; 2003-902907/82.
XX
PT Identifying a class I major histocompatibility complex (MHC)-binding
PT fragment of a polypeptide comprises isolating an MHC molecule, eluting
PT the peptide from the molecule, and identifying the peptide as a
PT polypeptide fragment.
XX
PS Claim 10; SEQ ID NO 7; 98pp; English.
XX
CC The present invention describes a method for identifying a class I major
CC histocompatibility complex (MHC)-binding fragment of a polypeptide by
CC isolating from the tissue/cell line a class I MHC molecule bound to a
CC peptide, where the peptide is a class I MHC-binding fragment of the
CC polypeptide, eluting the peptide from the class I MHC molecule, and
CC identifying the peptide as a fragment of the polypeptide. A class I MHC-
CC binding fragment has cytostatic activity. Compositions and methods from
CC the present invention can be used in diagnosing, preventing or treating
CC cancer. The method may also be used in identifying peptides involved in
CC the pathogenesis of or protection from diseases associated with
CC expression of class I MHC molecules. The present sequence represents a
CC human carcinoembryonic antigen (CEA) epitope peptide, which is used in
CC the exemplification of the present invention.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 41; DB 7; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|||||||||
Db 1 IMIGVLVGV 9

RESULT 13
AAW70047
ID AAW70047 standard; peptide; 10 AA.
XX
AC AAW70047;
XX
DT 22-OCT-1998 (first entry)
XX
DE CEA derived HLA-A2.1 binding peptide 4 (residues 690-699).
XX
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9833888-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US001959.
XX
PR 31-JAN-1997; 97US-0036696P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Tsai V, Southwood S, Sidney J, Sette A, Celis E;
XX
DR WPI; 1998-437445/37.
XX
PT Production of antigen-specific cytotoxic T cells - by incubating
PT immunogenic peptide(s) from antigen that binds class I major
PT histocompatibility complex molecules with pre-treated antigen presenting
PT cells.
XX
PS Example 6; Page 75; 104pp; English.
XX
CC Sequences shown in AAW70044 to AAW70052 represent peptides derived from
CC carcinoembryonic antigen (CEA). The peptides can bind to a human
CC leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC vitro. The method comprises contacting immunogenic peptides from an
CC antigen that binds class I major histocompatibility complex (MHC)
CC molecules with antigen presenting cells (APCs) pretreated with
CC pretreatment growth factors, and incubating the APCs with purified CD8
CC cells in the presence of at least 2 incubation growth factors, thereby
CC producing antigen-specific CTLs. A method for specifically killing target
CC cells in a human patient is also provided which comprises obtaining a
CC fluid sample containing CTLs from a patient, contacting the cytotoxic T
CC cells with APCs pretreated with pre-treatment growth factors, where the
CC APCs comprise class I MHC molecules. The pretreated APCs are incubated
CC with the cytotoxic growth factors, thereby producing activated CTLs which
CC are contacted with a carrier to form a composition. The composition can
CC then be administered to the patient. The activated CTLs can be used for
CC treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 41; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|||||||||
Db 2 IMIGVLVGV 10

RESULT 14
AAW70051
ID AAW70051 standard; peptide; 10 AA.
XX
AC AAW70051;
XX
DT 22-OCT-1998 (first entry)
XX
DE CEA derived HLA-A2.1 binding peptide 8 (residues 691-700).
XX
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW human leukocyte antigen; HLA; tumour associated antigen; cancer;

KW antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9833888-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US001959.
XX
PR 31-JAN-1997; 97US-0036696P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Tsai V, Southwood S, Sidney J, Sette A, Celis E;
XX
DR WPI; 1998-437445/37.
XX
PT Production of antigen-specific cytotoxic T cells - by incubating
PT immunogenic peptide(s) from antigen that binds class I major
PT histocompatibility complex molecules with pre-treated antigen presenting
PT cells.
XX
PS Example 6; Page 75; 104pp; English.
XX
CC Sequences shown in AAW70044 to AAW70052 represent peptides derived from
CC carcinoembryonic antigen (CEA). The peptides can bind to a human
CC leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC vitro. The method comprises contacting immunogenic peptides from an
CC antigen that binds class I major histocompatibility complex (MHC)
CC molecules with antigen presenting cells (APCs) pretreated with
CC pretreatment growth factors, and incubating the APCs with purified CD8
CC cells in the presence of at least 2 incubation growth factors, thereby
CC producing antigen-specific CTLs. A method for specifically killing target
CC cells in a human patient is also provided which comprises obtaining a
CC fluid sample containing CTLs from a patient, contacting the cytotoxic T
CC cells with APCs pretreated with pre-treatment growth factors, where the
CC APCs comprise class I MHC molecules. The pretreated APCs are incubated
CC with the cytotoxic growth factors, thereby producing activated CTLs which
CC are contacted with a carrier to form a composition. The composition can
CC then be administered to the patient. The activated CTLs can be used for
CC treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 41; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|||||||||
Db 1 IMIGVLVGV 9

RESULT 15
AAY47672
ID AAY47672 standard; peptide; 10 AA.
XX
AC AAY47672;
XX
DT 01-DEC-1999 (first entry)
XX
DE Immunogenic peptide having a human leukocyte antigen binding motif #2283.
XX
KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW immune response; T cell activation; major histocompatibility complex;
KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW vaccine; immunisation.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9945954-A1.
XX
PD 16-SEP-1999.
XX
PF 13-MAR-1998; 98WO-US005039.
XX
PR 13-MAR-1998; 98WO-US005039.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Kubo RT, Sidney J, Celis E, Grey HM, Southwood S;
XX
DR WPI; 1999-551214/46.
XX
PT New immunogenic peptides with HLA binding motif, useful in treatment and
PT diagnosis of cancers and viral diseases.
XX
PS Claim 1; Page 119; 150pp; English.
XX
CC AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
CC having a human major histocompatibility complex (MHC) Class I (also known
CC as human leukocyte antigen (HLA)) binding motif. The immunogenic peptides
CC can bind to a specific HLA allele (i.e. HLA-A subtypes HLA-A2.1, A1, A3.2
CC or A24.1 or HLA-B or C) and induce a cytotoxic T cell response against
CC the antigen from which the peptide is derived. Cytotoxic T lymphocytes
CC (CTLs) which destroy antigen-bearing cells are normally induced by an
CC antigen in the form of a peptide fragment bound to a HLA molecule, rather
CC than the intact foreign antigen itself, and are particularly important in
CC tumour rejection and in fighting viral infections. The peptides are
CC therefore useful therapeutically to treat or prevent viral infections and
CC cancers in mammals (especially humans) e.g. prostate cancer, hepatitis B
CC and C, AIDS, and renal carcinoma. They can be administered as vaccines to
CC elicit an immune response in individuals susceptible or otherwise at risk
CC of viral infection or cancer, or used to treat chronic or acute
CC conditions. They are also useful diagnostically, and can be used to
CC induce a cytotoxic T cell response, by contacting a cytotoxic T cell with
CC the peptide e.g. to produce CTLs ex vivo for infusion back into a
CC patient. The polynucleotides encoding the immunogenic peptides are also
CC useful therapeutically and for immunisation as above
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 41; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|||||||||
Db 1 IMIGVLVGV 9

Search completed: August 6, 2004, 08:34:13
Job time : 54 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 6, 2004, 08:32:45 ; Search time 16 Seconds
(without alignments)
54.108 Million cell updates/sec

Title: US-09-458-302B-193
Perfect score: 41
Sequence: 1 IMIGVLVGV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	702	2	A36319	carcinoembryonic a
2	38	92.7	75	2	A72733	hypothetical prote
3	37	90.2	447	2	AI0606	probable transport
4	36	87.8	357	2	S28058	serotonin receptor
5	36	87.8	357	2	I37107	5-HT5A serotonin r
6	36	87.8	357	2	B47472	5-hydroxytryptamin
7	36	87.8	553	2	T48486	hypothetical prote
8	35	85.4	95	2	AH2911	hypothetical prote
9	35	85.4	192	2	C69174	conserved hypothet
10	35	85.4	369	2	S38744	serotonin receptor
11	35	85.4	370	2	I48231	serotonin receptor
12	35	85.4	428	2	B83420	probable two-compo
13	34	82.9	58	2	AD2390	hypothetical prote
14	34	82.9	203	2	A97309	probable membrane
15	34	82.9	235	2	A44233	NADH2 dehydrogenas
16	34	82.9	265	2	A55811	carcinoembryonic a
17	34	82.9	327	2	D90530	hypothetical prote
18	34	82.9	415	2	AC1330	integral membrane
19	34	82.9	429	2	A56265	uracil transport p
20	34	82.9	429	2	G91048	uracil transport E
21	34	82.9	429	2	C85893	uracil transport [
22	34	82.9	447	2	H85596	hypothetical prote
23	34	82.9	447	2	D90746	probable transport
24	34	82.9	506	2	F83369	hypothetical prote
25	34	82.9	602	2	H88215	protein B0495.4 [i
26	34	82.9	633	2	S69734	hypothetical prote
27	34	82.9	634	2	T33528	hypothetical prote
28	34	82.9	805	2	A75014	hypothetical prote
29	34	82.9	890	2	T21000	hypothetical prote
30	33	80.5	41	2	S38918	hypothetical prote
31	33	80.5	89	2	T35348	probable membrane
32	33	80.5	116	2	S70489	photosystem II pro
33	33	80.5	230	2	T39598	hypothetical serin
34	33	80.5	250	2	C70823	hypothetical prote
35	33	80.5	279	2	AH1882	nitrate transport
36	33	80.5	289	2	E71498	probable yopt tran
37	33	80.5	294	2	AB1388	cell-division prot
38	33	80.5	294	2	AD1763	cell-division prot
39	33	80.5	297	2	A83049	hypothetical prote
40	33	80.5	312	2	G70114	phosphatidate cyti
41	33	80.5	370	2	E71221	hypothetical prote
42	33	80.5	372	1	D69442	conserved hypothet
43	33	80.5	419	2	C95091	conserved hypothet
44	33	80.5	419	2	G97958	conserved hypothet
45	33	80.5	473	2	B84143	Na+/H+ antiporter

ALIGNMENTS

RESULT 1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 31-Jan-2000
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098; A2
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shively,
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its pro
A;Reference number: A36319; MUID:90258861; PMID:2342461
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-Thr
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for human
A;Reference number: A27773; MUID:88038876; PMID:3670312
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and
A;Reference number: A31037; MUID:89122014; PMID:3220478
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA se
A;Reference number: A25845; MUID:87128144; PMID:3814146
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcinoe
A;Reference number: S31737

```
A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339; PMID:1427854
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118; PMID:2286372
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247; PMID:3033671
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349; PMID:3632664
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257; PMID:2372297
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form

  Query Match            100.0%;  Score 41;  DB 2;  Length 702;
  Best Local Similarity  100.0%;  Pred. No. 6.9;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db        691 IMIGVLVGV 699
```

```
RESULT 2
A72733
hypothetical protein APES012 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: A72733
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: A72733
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-75 <KAW>
A;Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BAA79357.1; PID:d1043143; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APES012

  Query Match            92.7%;  Score 38;  DB 2;  Length 75;
  Best Local Similarity  66.7%;  Pred. No. 2.9;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ||:|||:|:
Db         30 IMVGVLIGI 38
```

```
RESULT 3
AI0606
probable transport protein STY0917 [imported] - Salmonella enterica subsp. enterica sero
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AI0606
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AI0606
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-447 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05323.1; PID:g16502087; GSPDB:GN00176
C;Genetics:
A;Gene: STY0917

  Query Match            90.2%;  Score 37;  DB 2;  Length 447;
  Best Local Similarity  77.8%;  Pred. No. 24;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |:||:||||
Db         24 ILIGILVGV 32
```

```
RESULT 4
S28058
serotonin receptor 5 - mouse
N;Alternate names: 5-hydroxytryptamine 5 receptor (5HTR-5)
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999
C;Accession: S28058
R;Plassat, J.L.; Boschert, U.; Amlaiky, N.; Hen, R.
EMBO J. 11, 4779-4786, 1992
A;Title: The mouse 5HT5 receptor reveals a remarkable heterogeneity within the 5HT1D rec
A;Reference number: S28058; MUID:93099851; PMID:1464308
```

A;Accession: S28058
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-357 <PLA>
A;Cross-references: EMBL:Z18278
C;Superfamily: octopamine receptor type I
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; transme

Query Match 87.8%; Score 36; DB 2; Length 357;
Best Local Similarity 55.6%; Pred. No. 30;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
:|:|:|:||
Db 285 LMVGILIGV 293

RESULT 5
I37107
5-HT5A serotonin receptor - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I37107
R;Rees, S.; den Daas, I.; Foord, S.; Goodson, S.; Bull, D.; Kilpatrick, G.; Lee, M.
FEBS Lett. 355, 242-246, 1994
A;Title: Cloning and characterisation of the human 5-HT5A serotonin receptor.
A;Reference number: I37107; MUID:95080386; PMID:7988681
A;Accession: I37107
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <RES>
A;Cross-references: EMBL:X81411; NID:g541776; PIDN:CAA57168.1; PID:g784990
C;Genetics:
A;Gene: 5-HT5A
A;Introns: 247/3
C;Superfamily: octopamine receptor type I

Query Match 87.8%; Score 36; DB 2; Length 357;
Best Local Similarity 55.6%; Pred. No. 30;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
:|:|:|:||
Db 285 LMVGILIGV 293

RESULT 6
B47472
5-hydroxytryptamine 5 alpha receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: B47472
R;Erlander, M.G.; Lovenberg, T.W.; Baron, B.M.; de Lecea, L.; Danielson, P.E.; Racke, M.
Proc. Natl. Acad. Sci. U.S.A. 90, 3452-3456, 1993
A;Title: Two members of a distinct subfamily of 5-hydroxytryptamine receptors differenti
A;Reference number: A47472; MUID:93234515; PMID:7682702
A;Accession: B47472
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-357 <ERL>
A;Cross-references: GB:L10072; NID:g310072; PIDN:AAA40615.1; PID:g310073
A;Experimental source: hypothalamus
A;Note: sequence extracted from NCBI backbone (NCBIN:129674, NCBIP:129677)
C;Superfamily: octopamine receptor type I
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match 87.8%; Score 36; DB 2; Length 357;
Best Local Similarity 55.6%; Pred. No. 30;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
:|:|:|:||
Db 285 LMVGILIGV 293

RESULT 7
T48486
hypothetical protein T28J14.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48486
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; L
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T48486
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-553 <BEV>
A;Cross-references: EMBL:AL163652
A;Experimental source: cultivar Columbia; BAC clone T28J14
C;Genetics:
A;Map position: 5
A;Introns: 37/1; 80/2; 104/2; 280/1; 344/3; 387/1
A;Note: T28J14.90

Query Match 87.8%; Score 36; DB 2; Length 553;
Best Local Similarity 77.8%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|::||||||
Db 232 IIVGVLVGV 240

RESULT 8
AH2911
hypothetical protein Atu2729 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AH2911
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AH2911
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-95 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43710.1; PID:g17741239; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2729
A;Map position: circular chromosome

Query Match 85.4%; Score 35; DB 2; Length 95;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
||:|||:|:
Db 52 IMLGVLLGI 60

RESULT 9
C69174
conserved hypothetical protein MTH561 - Methanobacterium thermoautotrophicum (strain De
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
C;Accession: C69174
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: C69174
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-192 <MTH>
A;Cross-references: GB:AE000839; GB:AE000666; NID:g2621637; PIDN:AAB85067.1; PID:g262163
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH561
C;Superfamily: conserved hypothetical protein MJ0645

Query Match 85.4%; Score 35; DB 2; Length 192;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
::||||:||
Db 175 VVIGVLIGV 183

RESULT 10
S38744
serotonin receptor 5B - rat
N;Alternate names: 5-hydroxytryptamine receptor 5B (5-HTR5B)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C;Accession: S38744; A47472
R;Wisden, W.; Parker, E.M.; Mahle, C.D.; Grisel, D.A.; Nowak, H.P.; Yocca, F.D.; Felder,
FEBS Lett. 333, 25-31, 1993
A;Title: Cloning and characterization of the rat 5-HT(5B) receptor. Evidence that the 5-
A;Reference number: S38744; MUID:94039744; PMID:8224165
A;Accession: S38744
A;Molecule type: mRNA
A;Residues: 1-369 <WIS>
R;Erlander, M.G.; Lovenberg, T.W.; Baron, B.M.; de Lecea, L.; Danielson, P.E.; Racke, M.
Proc. Natl. Acad. Sci. U.S.A. 90, 3452-3456, 1993
A;Title: Two members of a distinct subfamily of 5-hydroxytryptamine receptors differenti
A;Reference number: A47472; MUID:93234515; PMID:7682702
A;Accession: A47472
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-176,'I',177-369 <ERL>
A;Cross-references: GB:L10073; NID:g310074; PIDN:AAA40616.1; PID:g310075
A;Experimental source: hypothalamus
A;Note: sequence extracted from NCBI backbone (NCBIN:129665, NCBIP:129668)
C;Superfamily: octopamine receptor type I
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match 85.4%; Score 35; DB 2; Length 369;
Best Local Similarity 55.6%; Pred. No. 47;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
:|:|:|:||
Db 297 MMVGILIGV 305

RESULT 11
I48231
serotonin receptor 5B - mouse
N;Alternate names: 5-hydroxytryptamine 5B receptor (5HTR-5C)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I48231
R;Matthes, H.; Boschert, U.; Amlaiky, N.; Grailhe, R.; Plassat, J.L.; Muscatelli, F.; Ma
Mol. Pharmacol. 43, 313-319, 1993
A;Title: Mouse 5-hydroxytryptamine5A and 5-hydroxytryptamine5B receptors define a new fa
A;Reference number: I48231; MUID:93196607; PMID:8450829
A;Accession: I48231
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-370 <RES>
A;Cross-references: EMBL:X69867; NID:g288735; PIDN:CAA49501.1; PID:g288736
C;Superfamily: octopamine receptor type I
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; transme

Query Match 85.4%; Score 35; DB 2; Length 370;
Best Local Similarity 55.6%; Pred. No. 47;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
:|:|:|:||
Db 298 MMVGILIGV 306

RESULT 12
B83420
probable two-component sensor PA1798 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83420
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83420
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <STO>
A;Cross-references: GB:AE004606; GB:AE004091; NID:g9947780; PIDN:AAG05187.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1798

Query Match 85.4%; Score 35; DB 2; Length 428;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|::|||||:
Db 140 ILLGVLVGI 148

RESULT 13
AD2390
hypothetical protein asr4676 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AD2390
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2390
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-58 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB76375.1; PID:g17133813; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: asr4676

Query Match 82.9%; Score 34; DB 2; Length 58;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|::||:|||
Db 20 IVVGVIVGV 28

```
RESULT 14
A97309
probable membrane protein [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: A97309
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A97309
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-203 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81260.1; PID:g15026409; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3328

  Query Match             82.9%;  Score 34;  DB 2;  Length 203;
  Best Local Similarity   77.8%;  Pred. No. 41;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||| :|||
Db         12 IMIGCIVGV 20


RESULT 15
A44233
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - fall armyworm mitochondrion (fra
N;Alternate names: NADH-ubiquinone oxidoreductase chain 1
C;Species: mitochondrion Spodoptera frugiperda (fall armyworm)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 03-Jun-2002
C;Accession: A44233
R;Pashley, D.P.; Ke, L.D.
Mol. Biol. Evol. 9, 1061-1075, 1992
A;Title: Sequence evolution in mitochondrial ribosomal and ND-1 genes in lepidoptera: im
A;Reference number: A44233; MUID:93061985; PMID:1435234
A;Accession: A44233
A;Molecule type: DNA
A;Residues: 1-235 <PAS>
A;Cross-references: GB:M76713; NID:g343352; PIDN:AAA32079.1; PID:g552886
A;Note: sequence extracted from NCBI backbone (NCBIP:118938)
C;Genetics:
A;Gene: ND-1
A;Genome: mitochondrion
A;Start codon: ATA
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 1
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

  Query Match             82.9%;  Score 34;  DB 2;  Length 235;
  Best Local Similarity   66.7%;  Pred. No. 47;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ::||:||||
Db         14 LIIGILVGV 22


Search completed: August  6, 2004, 08:35:27
Job time : 18 secs
```

This Page Blank (uspto)

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 6, 2004, 08:32:42 ; Search time 13 Seconds
(without alignments)
36.049 Million cell updates/sec

Title: US-09-458-302B-193
Perfect score: 41
Sequence: 1 IMIGVLVGV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	702	1	CEA5_HUMAN	P06731 homo sapien
2	36	87.8	357	1	5H5A_HUMAN	P47898 homo sapien
3	36	87.8	357	1	5H5A_MOUSE	P30966 mus musculu
4	36	87.8	357	1	5H5A_RAT	P35364 rattus norv
5	35	85.4	370	1	5H5B_MOUSE	P31387 mus musculu
6	35	85.4	370	1	5H5B_RAT	P35365 rattus norv
7	34	82.9	238	1	RNFE_AZOVI	Q9f5y1 azotobacter
8	34	82.9	265	1	CEA7_HUMAN	Q14002 homo sapien
9	34	82.9	429	1	URAA_ECOLI	P33780 escherichia
10	34	82.9	633	1	NAH2_YEAST	Q04121 saccharomyc
11	34	82.9	685	1	ST41_ARATH	Q9fy46 arabidopsis
12	33	80.5	78	1	YK61_LACPL	Q88vi8 lactobacill
13	33	80.5	333	1	PTHB_ERWAM	O32522 erwinia amy
14	33	80.5	660	1	Y390_MYCGE	Q49430 mycoplasma
15	33	80.5	677	1	ST42_ARATH	Q8gyh8 arabidopsis
16	32	78.0	73	1	Y0I0_BACCR	Q812y8 bacillus ce
17	32	78.0	73	1	Y1E2_BACAA	Q81y17 bacillus an
18	32	78.0	174	1	NU6M_HYLLA	Q95710 hylobates l
19	32	78.0	205	1	YO01_BPHP1	P51700 bacteriopha
20	32	78.0	241	1	RNFE_PSEST	Q9evn2 pseudomonas
21	32	78.0	241	1	Y513_METJA	Q57933 methanococc
22	32	78.0	245	1	COBS_PSEAE	Q9i463 pseudomonas
23	32	78.0	252	1	CEA3_HUMAN	P40198 homo sapien
24	32	78.0	278	1	UPK_SULSO	Q97x94 sulfolobus
25	32	78.0	291	1	UPK2_STRCO	Q9k407 streptomyce
26	32	78.0	317	1	NU1M_DICDI	Q37313 dictyosteli
27	32	78.0	319	1	PTHB_ECOLI	P56580 escherichia
28	32	78.0	414	1	CP51_ISSOR	Q02315 issatchenki
29	32	78.0	421	1	HEMA_CVBF	P33468 bovine coro
30	32	78.0	424	1	HEMA_CVBEN	P59711 bovine coro
31	32	78.0	424	1	HEMA_CVBG9	Q66165 bovine coro
32	32	78.0	424	1	HEMA_CVBL9	P59710 bovine coro
33	32	78.0	424	1	HEMA_CVBLS	Q9qar6 bovine coro
34	32	78.0	424	1	HEMA_CVBLU	Q8v437 bovine coro
35	32	78.0	424	1	HEMA_CVBLY	P31613 bovine coro
36	32	78.0	424	1	HEMA_CVBM	P15776 bovine coro
37	32	78.0	424	1	HEMA_CVBQ	P59709 bovine coro
38	32	78.0	424	1	HEMA_CVHOC	P30215 human coron
39	32	78.0	424	1	HEMA_CVP67	Q8bb26 porcine hem
40	32	78.0	424	1	HEMA_CVPIA	Q8jsp9 porcine hem
41	32	78.0	428	1	HEMA_CVMA5	P31615 murine coro
42	32	78.0	430	1	HEMA_CVMS	P31614 murine coro
43	32	78.0	431	1	HEMA_CVMDV	O92367 murine coro
44	32	78.0	438	1	PBUX_BACSU	P42086 bacillus su
45	32	78.0	439	1	HEMA_CVPV	O91262 puffinosis

ALIGNMENTS

RESULT 1
CEA5_HUMAN
ID CEA5_HUMAN STANDARD; PRT; 702 AA.
AC P06731;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66e
DE antigen).
GN CEACAM5 OR CEA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258861; PubMed=2342461;
RA Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT of its promoter indicates a region conveying cell type-specific
RT expression.";
RL Mol. Cell. Biol. 10:2738-2748(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88038876; PubMed=3670312;
RA Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT "Isolation and characterization of full-length functional cDNA clones
RT for human carcinoembryonic antigen.";
RL Mol. Cell. Biol. 7:3221-3230(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89122014; PubMed=3220478;
RA Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT "Carcinoembryonic antigen family: characterization of cDNAs coding
RT for NCA and CEA and suggestion of nonrandom sequence variation in
RT their conserved loop-domains.";
RL Genomics 3:59-66(1988).
RN [4]
RP SEQUENCE OF 5-702 FROM N.A.
RX MEDLINE=87128144; PubMed=3814146;
RA Oikawa S., Nakazato H., Kosaki G.;
RT "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT from cDNA sequence.";
RL Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN [5]
RP SEQUENCE OF 331-702 FROM N.A.
RX MEDLINE=87204247; PubMed=3033671;
RA Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT "Isolation and characterization of cDNA clones encoding the human
RT carcinoembryonic antigen reveal a highly conserved repeating
RT structure.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- TISSUE SPECIFICITY: Found in adenocarcinomas of endodermally

```
CC       derived digestive system epithelium and fetal colon.
CC   -!- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 kDa
CC       COMPRISING 60% CARBOHYDRATE.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd66e.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   PDB; 1E07; 04-JUL-00.
DR   Genew; HGNC:1817; CEACAM5.
DR   MIM; 114890; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 6.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   Immunoglobulin domain; Glycoprotein; Lipoprotein; GPI-anchor;
KW   Membrane; Signal; Repeat; 3D-structure.
FT   SIGNAL       1    34
FT   CHAIN       35   685       CARCINOEMBRYONIC ANTIGEN-RELATED CELL
FT                               ADHESION MOLECULE 5.
FT   PROPEP     686   702       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN      35   144       IG-LIKE 1.
FT   DOMAIN     146   237       IG-LIKE 2.
FT   DOMAIN     238   322       IG-LIKE 3.
FT   DOMAIN     324   415       IG-LIKE 4.
FT   DOMAIN     416   498       IG-LIKE 5.
FT   DOMAIN     502   593       IG-LIKE 6.
FT   DOMAIN     594   677       IG-LIKE 7.
FT   LIPID      685   685       GPI-anchor amidated alanine (Potential).
FT   CARBOHYD   104   104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   115   115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   152   152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   182   182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   197   197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   204   204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   208   208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   246   246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   256   256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   274   274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   288   288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   292   292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   309   309       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   330   330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   351   351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   360   360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   375   375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   432   432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   466   466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   480   480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   508   508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   529   529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   553   553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   560   560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   580   580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   612   612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   650   650       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   665   665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   320   320       MISSING (IN REF. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;

  Query Match             100.0%;  Score 41;  DB 1;  Length 702;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||||||
Db        691 IMIGVLVGV 699


RESULT 2
5H5A_HUMAN
ID   5H5A_HUMAN     STANDARD;      PRT;   357 AA.
AC   P47898;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   5-hydroxytryptamine 5A receptor (5-HT-5A) (Serotonin receptor)
DE   (5-HT-5).
GN   HTR5A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95080386; PubMed=7988681;
RA   Rees S., den Daas I., Foord S., Goodson S., Bull D., Kilpatrick G.,
RA   Lee M.;
RT   "Cloning and characterisation of the human 5-HT5A serotonin
RT   receptor.";
RL   FEBS Lett. 355:242-246(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Puhl H.L. III, Ikeda S.R., Aronstam R.S.;
RT   "cDNA clones of human proteins involved in signal transduction
RT   sequenced by the Guthrie cDNA resource center (www.cdna.org).";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This is one of the several different receptors for 5-
CC       hydroxytryptamine (serotonin), a biogenic hormone that functions
CC       as a neurotransmitter, a hormone, and a mitogen. The activity of
CC       this receptor is mediated by G proteins.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC       STRONGEST TO THE OTHER 5HT-5 SUBTYPE RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X81411; CAA57168.1; -.
DR   EMBL; X81412; CAA57168.1; JOINED.
DR   EMBL; AF498985; AAM21132.1; -.
DR   PIR; I37107; I37107.
DR   Genew; HGNC:5300; HTR5A.
DR   MIM; 601305; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004993; F:serotonin receptor activity; TAS.
```

```
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     40       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     41     63       1 (POTENTIAL).
FT   DOMAIN       64     78       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     79     99       2 (POTENTIAL).
FT   DOMAIN      100    115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    116    137       3 (POTENTIAL).
FT   DOMAIN      138    158       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    159    181       4 (POTENTIAL).
FT   DOMAIN      182    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    219       5 (POTENTIAL).
FT   DOMAIN      220    282       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    283    303       6 (POTENTIAL).
FT   DOMAIN      304    320       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    321    341       7 (POTENTIAL).
FT   DOMAIN      342    357       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      6      6       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    120    192       BY SIMILARITY.
SQ   SEQUENCE   357 AA;  40255 MW;  92F8A78C69169790 CRC64;

  Query Match             87.8%;  Score 36;  DB 1;  Length 357;
  Best Local Similarity   55.6%;  Pred. No. 19;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|:|:||
Db        285 LMVGILIGV 293


RESULT 3
5H5A_MOUSE
ID   5H5A_MOUSE     STANDARD;      PRT;   357 AA.
AC   P30966;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   5-hydroxytryptamine 5A receptor (5-HT-5A) (Serotonin receptor)
DE   (5-HT-5).
GN   HTR5A OR 5HT5A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93099851; PubMed=1464308;
RA   Plassat J.-L., Boschert U., Amlaiky N., Hen R.;
RT   "The mouse 5HT5 receptor reveals a remarkable heterogeneity within
RT   the 5HT1D receptor family.";
RL   EMBO J. 11:4779-4786(1992).
RN   [2]
RP   CHARACTERIZATION.
RC   TISSUE=Brain;
RX   MEDLINE=93196607; PubMed=8450829;
RA   Matthes H., Boschert U., Amlaiky N., Grailhe R., Plassat J.-L.,
RA   Muscatelli F., Mattei M.-G., Hen R.;
RT   "Mouse 5-hydroxytryptamine5A and 5-hydroxytryptamine5B receptors
RT   define a new family of serotonin receptors: cloning, functional
RT   expression, and chromosomal localization.";
RL   Mol. Pharmacol. 43:313-319(1993).
CC   -!- FUNCTION: This is one of the several different receptors for 5-
CC       hydroxytryptamine (serotonin), a biogenic hormone that functions
CC       as a neurotransmitter, a hormone, and a mitogen. The activity of
CC       this receptor is mediated by G proteins.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in the central nervous
CC       system; in the cerebral cortex, hippocampus, habenula, olfactory
CC       bulb and granular layer of the cerebellum.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC       STRONGEST TO THE OTHER 5HT-5 SUBTYPE RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z18278; CAA79155.1; -.
DR   MGD; MGI:96283; Htr5a.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     40       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     41     63       1 (POTENTIAL).
FT   DOMAIN       64     78       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     79     99       2 (POTENTIAL).
FT   DOMAIN      100    115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    116    137       3 (POTENTIAL).
FT   DOMAIN      138    158       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    159    181       4 (POTENTIAL).
FT   DOMAIN      182    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    219       5 (POTENTIAL).
FT   DOMAIN      220    282       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    283    303       6 (POTENTIAL).
FT   DOMAIN      304    320       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    321    341       7 (POTENTIAL).
FT   DOMAIN      342    357       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      6      6       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    120    192       BY SIMILARITY.
SQ   SEQUENCE   357 AA;  40804 MW;  5F5D856AC477BFAC CRC64;

  Query Match             87.8%;  Score 36;  DB 1;  Length 357;
  Best Local Similarity   55.6%;  Pred. No. 19;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|:|:||
Db        285 LMVGILIGV 293


RESULT 4
5H5A_RAT
ID   5H5A_RAT       STANDARD;      PRT;   357 AA.
AC   P35364;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   5-hydroxytryptamine 5A receptor (5-HT-5A) (Serotonin receptor)
DE   (REC17).
GN   HTR5A OR 5HT5A.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=93234515; PubMed=7682702;
RA   Erlander M.G., Lovenberg T.W., Baron B.M., de Lecea L.,
RA   Danielson P.E., Racke M., Slone A.L., Siegel B.W., Foye P.E.,
```

```
RA   Cannon K., Burns J.E., Sutcliffe G.J.;
RT   "Two members of a distinct subfamily of 5-hydroxytryptamine receptors
RT   differentially expressed in rat brain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:3452-3456(1993).
CC   -!- FUNCTION: This is one of the several different receptors for 5-
CC       hydroxytryptamine (serotonin), a biogenic hormone that functions
CC       as a neurotransmitter, a hormone, and a mitogen. The activity of
CC       this receptor is mediated by G proteins.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Central nervous system.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC       STRONGEST TO THE OTHER 5HT-5 SUBTYPE RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L10072; AAA40615.1; -.
DR   PIR; B47472; B47472.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     40       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     41     63       1 (POTENTIAL).
FT   DOMAIN       64     78       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     79     99       2 (POTENTIAL).
FT   DOMAIN      100    115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    116    137       3 (POTENTIAL).
FT   DOMAIN      138    158       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    159    181       4 (POTENTIAL).
FT   DOMAIN      182    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    219       5 (POTENTIAL).
FT   DOMAIN      220    282       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    283    303       6 (POTENTIAL).
FT   DOMAIN      304    320       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    321    341       7 (POTENTIAL).
FT   DOMAIN      342    357       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      6      6       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    120    192       BY SIMILARITY.
SQ   SEQUENCE   357 AA;  40672 MW;  8C498A50C88408B5 CRC64;

  Query Match             87.8%;  Score 36;  DB 1;  Length 357;
  Best Local Similarity   55.6%;  Pred. No. 19;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|:|:||
Db        285 LMVGILIGV 293


RESULT 5
5H5B_MOUSE
ID   5H5B_MOUSE     STANDARD;      PRT;   370 AA.
AC   P31387;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   5-hydroxytryptamine 5B receptor (5-HT-5B) (Serotonin receptor).
GN   HTR5B OR 5HT5B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93196607; PubMed=8450829;
RA   Matthes H., Boschert U., Amlaiky N., Grailhe R., Plassat J.-L.,
RA   Muscatelli F., Mattei M.-G., Hen R.;
RT   "Mouse 5-hydroxytryptamine5A and 5-hydroxytryptamine5B receptors
RT   define a new family of serotonin receptors: cloning, functional
RT   expression, and chromosomal localization.";
RL   Mol. Pharmacol. 43:313-319(1993).
CC   -!- FUNCTION: This is one of the several different receptors for
CC       5-hydroxytryptamine (serotonin), a biogenic hormone that functions
CC       as a neurotransmitter, a hormone, and a mitogen. The activity of
CC       this receptor is mediated by G proteins. Probably involved in
CC       anxiety and depression.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in the central nervous
CC       system; in the hippocampus, habenula, and the doral raphe.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC       STRONGEST TO THE OTHER 5HT-5 SUBTYPE RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69867; CAA49501.1; -.
DR   PIR; I48231; I48231.
DR   MGD; MGI:96284; Htr5b.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     52       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     53     75       1 (POTENTIAL).
FT   DOMAIN       76     90       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     91    111       2 (POTENTIAL).
FT   DOMAIN      112    128       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    129    150       3 (POTENTIAL).
FT   DOMAIN      151    171       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    172    194       4 (POTENTIAL).
FT   DOMAIN      195    211       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    212    232       5 (POTENTIAL).
FT   DOMAIN      233    295       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    296    316       6 (POTENTIAL).
FT   DOMAIN      317    333       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    334    354       7 (POTENTIAL).
FT   DOMAIN      355    370       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      5      5       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    127    205       BY SIMILARITY.
SQ   SEQUENCE   370 AA;  41201 MW;  0553C62B12DAAD84 CRC64;

  Query Match             85.4%;  Score 35;  DB 1;  Length 370;
  Best Local Similarity   55.6%;  Pred. No. 29;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|:|:||
Db        298 MMVGILIGV 306


RESULT 6
5H5B_RAT
ID   5H5B_RAT       STANDARD;      PRT;   370 AA.
AC   P35365;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
```

```
DE   5-hydroxytryptamine 5B receptor (5-HT-5B) (Serotonin receptor) (MR22).
GN   HTR5B OR 5HT5B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=93234515; PubMed=7682702;
RA   Erlander M.G., Lovenberg T.W., Baron B.M., de Lecea L.,
RA   Danielson P.E., Racke M., Slone A.L., Siegel B.W., Foye P.E.,
RA   Cannon K., Burns J.E., Sutcliffe G.J.;
RT   "Two members of a distinct subfamily of 5-hydroxytryptamine receptors
RT   differentially expressed in rat brain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:3452-3456(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94039744; PubMed=8224165;
RA   Wisden W., Parker E.M., Mahle C.D., Grisel D.A., Nowak H.P.,
RA   Yocca F.D., Felder C.C., Seeburg P.H., Voigt M.M.;
RT   "Cloning and characterization of the rat 5-HT5B receptor. Evidence
RT   that the 5-HT5B receptor couples to a G protein in mammalian cell
RT   membranes.";
RL   FEBS Lett. 333:25-31(1993).
CC   -!- FUNCTION: This is one of the several different receptors for 5-
CC       hydroxytryptamine (serotonin), a biogenic hormone that functions
CC       as a neurotransmitter, a hormone, and a mitogen. The activity of
CC       this receptor is mediated by G proteins. Probably involved in
CC       anxiety and depression.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Brain; in the CA1 region of hippocampus, the
CC       medial habenula, and raphe nuclei.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC       STRONGEST TO THE OTHER 5HT-5 SUBTYPE RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L10073; AAA40616.1; -.
DR   PIR; S38744; S38744.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     52       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     53     75       1 (POTENTIAL).
FT   DOMAIN       76     90       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     91    111       2 (POTENTIAL).
FT   DOMAIN      112    128       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    129    150       3 (POTENTIAL).
FT   DOMAIN      151    171       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    172    194       4 (POTENTIAL).
FT   DOMAIN      195    211       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    212    232       5 (POTENTIAL).
FT   DOMAIN      233    295       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    296    316       6 (POTENTIAL).
FT   DOMAIN      317    333       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    334    354       7 (POTENTIAL).
FT   DOMAIN      355    370       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      5      5       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    127    205       BY SIMILARITY.
SQ   SEQUENCE   370 AA;  41122 MW;  8EC5F789BFD647E5 CRC64;

  Query Match             85.4%;  Score 35;  DB 1;  Length 370;
  Best Local Similarity   55.6%;  Pred. No. 29;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|:|:||
Db        298 MMVGILIGV 306


RESULT 7
RNFE_AZOVI
ID   RNFE_AZOVI     STANDARD;      PRT;   238 AA.
AC   Q9F5Y1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Electron transport complex protein rnfE (Nitrogen fixation protein
DE   rnfE).
GN   RNFE.
OS   Azotobacter vinelandii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Azotobacter.
OX   NCBI_TaxID=354;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DJ;
RA   Rubio L.M., Rangaraj P., Roberts G.P., Ludden P.W.;
RT   "Cloning and mutational analysis of the Azotobacter vinelandii gene
RT   encoding the dinitrogenase gamma subunit.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Required for nitrogen fixation. May be part of a
CC       membrane complex functioning as an intermediate in the electron
CC       transport to nitrogenase (By similarity).
CC   -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC       rnfD, rnfE and rnfG (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: Belongs to the nqrDE/rnfAE family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF302049; AAG29820.1; -.
DR   HAMAP; MF_00478; -; 1.
DR   InterPro; IPR003667; Rnf_Nqr.
DR   Pfam; PF02508; Rnf-Nqr; 1.
KW   Nitrogen fixation; Electron transport; Transmembrane; Inner membrane.
FT   TRANSMEM     41     63       POTENTIAL.
FT   TRANSMEM     84    104       POTENTIAL.
FT   TRANSMEM    106    126       POTENTIAL.
FT   TRANSMEM    141    161       POTENTIAL.
FT   TRANSMEM    195    215       POTENTIAL.
SQ   SEQUENCE   238 AA;  25527 MW;  5701ADD4D1D55734 CRC64;

  Query Match             82.9%;  Score 34;  DB 1;  Length 238;
  Best Local Similarity   66.7%;  Pred. No. 31;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :||||: ||
Db         84 VMIGVIAGV 92


RESULT 8
CEA7_HUMAN
ID   CEA7_HUMAN     STANDARD;      PRT;   265 AA.
AC   Q14002; O15148; O15149; Q9UPJ2;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
```

```
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Carcinoembryonic antigen-related cell adhesion molecule 7 precursor
DE   (Carcinoembryonic antigen CGM2).
GN   CEACAM7 OR CGM2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2A).
RX   MEDLINE=95105177; PubMed=7806520;
RA   Thompson J., Zimmermann W., Nollau P., Neumaier M., Weber-Arden J.,
RA   Schrewe H., Craig I., Willcocks T.;
RT   "CGM2, a member of the carcinoembryonic antigen gene family is down-
RT   regulated in colorectal carcinomas.";
RL   J. Biol. Chem. 269:32924-32931(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2A).
RC   TISSUE=Colon mucosa;
RX   MEDLINE=97280695; PubMed=9135022;
RA   Thompson J., Seitz M., Chastre E., Ditter M., Aldrian C., Gespach C.,
RA   Zimmermann W.;
RT   "Down-regulation of carcinoembryonic antigen family member 2
RT   expression is an early event in colorectal tumorigenesis.";
RL   Cancer Res. 57:1776-1784(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 2A AND 2B).
RA   Zhou G.Q.;
RT   "Two isoforms of CEA gene family member 2 (CGM2) mRNA are co-expressed
RT   in small and large intestine mucosa epithelium and in colorectal tumor
RT   cells.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS 2A AND 2B).
RA   Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA   Burkhart-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA   Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA   Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA   Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA   Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 2.5 Mb region in 19q13.2 containing a
RT   clustered CEA/PSG gene family.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=2a;
CC         IsoId=Q14002-1; Sequence=Displayed;
CC       Name=2b;
CC         IsoId=Q14002-2; Sequence=VSP_002488;
CC   -!- TISSUE SPECIFICITY: Strongly down-regulated in colonic
CC       adenocarcinomas.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31792; AAA66186.1; -.
DR   EMBL; X98311; CAA66955.1; -.
DR   EMBL; AF006622; AAB62924.1; -.
DR   EMBL; AF006623; AAB62925.1; -.
DR   EMBL; AC005797; AAC62825.1; -.
DR   EMBL; AC005797; AAC62826.1; -.
DR   PIR; A55811; A55811.
DR   Genew; HGNC:1819; CEACAM7.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0005886; C:plasma membrane; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   SMART; SM00409; IG; 2.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin domain; Antigen; Membrane; Signal; Glycoprotein;
KW   Lipoprotein; GPI-anchor; Repeat; Alternative splicing.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    242       CARCINOEMBRYONIC ANTIGEN-RELATED CELL
FT                                 ADHESION MOLECULE 7.
FT   PROPEP      243    265       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       35    142       IG-LIKE V-TYPE.
FT   DOMAIN      146    233       IG-LIKE C2-TYPE.
FT   DISULFID    168    216       PROBABLE.
FT   LIPID       242    242       GPI-anchor amidated serine (Potential).
FT   CARBOHYD     57     57       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    105    105       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    183    183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    198    198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    143    235       Missing (in isoform 2b).
FT                                 /FTId=VSP_002488.
FT   CONFLICT     40     40       V -> G (IN REF. 3).
FT   CONFLICT     71     71       E -> Q (IN REF. 3).
FT   CONFLICT    120    120       F -> I (IN REF. 3; AAB62925).
FT   CONFLICT    235    235       R -> C (IN REF. 3).
SQ   SEQUENCE   265 AA;  29379 MW;  B6B836F66BD10D3B CRC64;

  Query Match             82.9%;  Score 34;  DB 1;  Length 265;
  Best Local Similarity   77.8%;  Pred. No. 34;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |||||| |:
Db        254 IMIGVLAGM 262


RESULT 9
URAA_ECOLI
ID   URAA_ECOLI     STANDARD;      PRT;   429 AA.
AC   P33780;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Uracil permease (Uracil transporter).
GN   URAA OR B2497 OR Z3760 OR ECS3359.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=95238271; PubMed=7721693;
RA   Andersen P.S., Frees D., Fast R., Mygind B.;
RT   "Uracil uptake in Escherichia coli K-12: isolation of uraA mutants
RT   and cloning of the gene.";
RL   J. Bacteriol. 177:2008-2013(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
```

```
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: Transport of uracil in the cell.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X73586; CAA51992.1; -.
DR   EMBL; AE000336; AAC75550.1; -.
DR   EMBL; D90878; BAA16385.1; -.
DR   EMBL; AE005479; AAG57607.1; -.
DR   EMBL; AP002561; BAB36782.1; -.
DR   PIR; A56265; A56265.
DR   PIR; C85893; C85893.
DR   PIR; G91048; G91048.
DR   EcoGene; EG12129; uraA.
DR   InterPro; IPR006042; Xan_ur_permease.
DR   InterPro; IPR006043; Xant/urac/vitC.
DR   Pfam; PF00860; xan_ur_permease; 1.
DR   TIGRFAMs; TIGR00801; ncs2; 1.
DR   PROSITE; PS01116; XANTH_URACIL_PERMASE; 1.
KW   Transmembrane; Transport; Inner membrane; Complete proteome.
FT   TRANSMEM     29     49       POTENTIAL.
FT   TRANSMEM     65     85       POTENTIAL.
FT   TRANSMEM     88    108       POTENTIAL.
FT   TRANSMEM    127    147       POTENTIAL.
FT   TRANSMEM    159    179       POTENTIAL.
FT   TRANSMEM    182    202       POTENTIAL.
FT   TRANSMEM    228    248       POTENTIAL.
FT   TRANSMEM    300    320       POTENTIAL.
FT   TRANSMEM    325    345       POTENTIAL.
FT   TRANSMEM    366    386       POTENTIAL.
FT   TRANSMEM    392    412       POTENTIAL.
SQ   SEQUENCE   429 AA;  45060 MW;  18045190C960C674 CRC64;

  Query Match             82.9%;  Score 34;  DB 1;  Length 429;
  Best Local Similarity   87.5%;  Pred. No. 50;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVG 8
              |:||||||
Db        185 ILIGVLVG 192


RESULT 10
NAH2_YEAST
ID   NAH2_YEAST     STANDARD;      PRT;   633 AA.
AC   Q04121;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mitochondrial sodium/hydrogen exchanger (Mitochondrial Na(+)/H(+)
DE   exchanger).
GN   NHA2 OR NHX1 OR YDR456W OR D9461.40.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
RA   Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA   Hunicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA   Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA   Winant A., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98175963; PubMed=9507001;
RA   Numata M., Petrecca K., Lake N., Orlowski J.;
RT   "Identification of a mitochondrial Na+/H+ exchanger.";
RL   J. Biol. Chem. 273:6951-6959(1998).
CC   -!- FUNCTION: Electroneutral exchange of protons for Na(+) and K(+)
CC       across the mitochondrial inner membrane. Contributes to organellar
CC       volume and calcium homeostasis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial.
CC   -!- SIMILARITY: Belongs to the Na(+)/H(+) exchanger family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U33007; AAB64861.1; -.
DR   PIR; S69734; S69734.
DR   GermOnline; 140948; -.
DR   SGD; S0002864; NHX1.
DR   GO; GO:0005770; C:late endosome; IDA.
DR   GO; GO:0015077; F:monovalent inorganic cation transporter act. . .; IDA.
DR   GO; GO:0016197; P:endosome transport; IMP.
DR   GO; GO:0030004; P:monovalent inorganic cation homeostasis; IMP.
DR   GO; GO:0015672; P:monovalent inorganic cation transport; IMP.
DR   GO; GO:0007035; P:vacuolar acidification; IMP.
DR   InterPro; IPR006153; Na_H_porter.
DR   InterPro; IPR004709; NaH_exchang.
DR   Pfam; PF00999; Na_H_Exchanger; 1.
DR   PRINTS; PR01084; NAHEXCHNGR.
DR   TIGRFAMs; TIGR00840; b_cpa1; 1.
KW   Transmembrane; Transport; Antiport; Sodium transport; Mitochondrion.
```

```
FT   TRANSMEM     62     82       POTENTIAL.
FT   TRANSMEM     86    106       POTENTIAL.
FT   TRANSMEM    118    138       POTENTIAL.
FT   TRANSMEM    154    174       POTENTIAL.
FT   TRANSMEM    177    197       POTENTIAL.
FT   TRANSMEM    218    238       POTENTIAL.
FT   TRANSMEM    259    279       POTENTIAL.
FT   TRANSMEM    307    327       POTENTIAL.
FT   TRANSMEM    353    373       POTENTIAL.
FT   TRANSMEM    377    397       POTENTIAL.
FT   TRANSMEM    432    452       POTENTIAL.
FT   TRANSMEM    458    478       POTENTIAL.
SQ   SEQUENCE   633 AA;  70147 MW;  9B771ABDE41CEB0A CRC64;

  Query Match             82.9%;  Score 34;  DB 1;  Length 633;
  Best Local Similarity   55.6%;  Pred. No. 69;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ::||||:|:
Db        268 LLIGVLIGI 276


RESULT 11
ST41_ARATH
ID   ST41_ARATH     STANDARD;      PRT;   685 AA.
AC   Q9FY46; O22123; Q8GW68; Q9FNB8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Sulfate transporter 4.1, chloroplast precursor (AST82).
GN   SULTR4;1 OR AT5G13550 OR T6I14_80 OR MSH12.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Takahashi H., Asanuma W., Saito K.;
RT   "Cloning of an Arabidopsis cDNA encoding a chloroplast localizing
RT   sulphate transporter isoform.";
RL   J. Exp. Bot. 50:1713-1714(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016721; PubMed=11130714;
RA   Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA   Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA   Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA   Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA   Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA   Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA   Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA   Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA   Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA   Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,
RA   Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA   Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA   Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
RA   Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA   Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA   Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA   Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA   Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA   van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA   Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA   Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA   Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
RA   Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;
RT   "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:823-826(2000).
RN   [3]
RP   SEQUENCE OF 1-389 FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=98069011; PubMed=9405937;
RA   Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT   Sequence features of the regions of 1,044,062 bp covered by thirteen
RT   physically assigned P1 clones.";
RL   DNA Res. 4:291-300(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA   Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA   Hayashizaki Y., Shinozaki K.;
RT   "Arabidopsis thaliana full-length cDNA.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   INDUCTION.
RX   MEDLINE=20387013; PubMed=10929111;
RA   Takahashi H., Watanabe-Takahashi A., Smith F.W., Blake-Kalff M.,
RA   Hawkesford M.J., Saito K.;
RT   "The roles of three functional sulphate transporters involved in
RT   uptake and translocation of sulphate in Arabidopsis thaliana.";
RL   Plant J. 23:171-182(2000).
CC   -!- FUNCTION: H(+)/sulfate cotransporter that may play a role in the
CC       regulation of sulfate assimilation.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC       Chloroplast.
CC   -!- TISSUE SPECIFICITY: Expressed both in roots and leaves.
CC   -!- INDUCTION: By sulfate starvation in leaves.
CC   -!- SIMILARITY: Belongs to the SLC26A/SulP transporter (TC 2.A.53)
CC       family.
CC   -!- SIMILARITY: Contains 1 STAS domain.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to a stop
CC       codon in position 353 which was translated as Lys to extend the
CC       sequence.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB008782; BAA23424.1; -.
DR   EMBL; AL391710; CAC05432.1; -.
DR   EMBL; AB006704; BAB17026.1; -.
DR   EMBL; AK119047; BAC43623.1; ALT_TERM.
DR   InterPro; IPR002645; STAS.
DR   InterPro; IPR001902; Sulph_transpt.
DR   Pfam; PF01740; STAS; 1.
DR   Pfam; PF00916; Sulfate_transp; 1.
DR   TIGRFAMs; TIGR00815; sulP; 1.
DR   PROSITE; PS01130; SLC26A; 1.
DR   PROSITE; PS50801; STAS; 1.
KW   Transport; Symport; Sulfate transport; Transit peptide; Chloroplast;
KW   Transmembrane; Multigene family.
FT   TRANSIT       1     23       CHLOROPLAST (POTENTIAL).
FT   CHAIN        24    685       SULFATE TRANSPORTER 4.1.
FT   TRANSMEM     97    117       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    147    167       POTENTIAL.
FT   TRANSMEM    175    195       POTENTIAL.
FT   TRANSMEM    203    223       POTENTIAL.
FT   TRANSMEM    255    275       POTENTIAL.
FT   TRANSMEM    283    303       POTENTIAL.
FT   TRANSMEM    332    352       POTENTIAL.
FT   TRANSMEM    369    389       POTENTIAL.
```

```
FT   TRANSMEM     406     426       POTENTIAL.
FT   TRANSMEM     434     454       POTENTIAL.
FT   TRANSMEM     473     493       POTENTIAL.
FT   DOMAIN       518     642       STAS.
FT   CONFLICT     229     229       L -> Q (IN REF. 4).
FT   CONFLICT     344     344       A -> P (IN REF. 1).
FT   CONFLICT     368     368       E -> D (IN REF. 1).
SQ   SEQUENCE   685 AA;  75095 MW;  8C0087229BC39ADD CRC64;

  Query Match             82.9%;  Score 34;  DB 1;  Length 685;
  Best Local Similarity   88.9%;  Pred. No. 74;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              | |||||||
Db        477 IEIGVLVGV 485


RESULT 12
YK61_LACPL
ID   YK61_LACPL     STANDARD;      PRT;    78 AA.
AC   Q88VI8;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hypothetical UPF0154 protein lp_2061.
GN   LP_2061.
OS   Lactobacillus plantarum.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1590;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIMB 8826 / WCFS1;
RX   MEDLINE=22480296; PubMed=12566566;
RA   Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA   Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA   Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA   Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA   De Vos W.M., Siezen R.J.;
RT   "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
CC   -!- SIMILARITY: Belongs to the UPF0154 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL935258; CAD64433.1; -.
DR   HAMAP; MF_00363; -; 1.
DR   InterPro; IPR005359; UPF0154.
DR   Pfam; PF03672; UPF0154; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      5      27       Potential.
SQ   SEQUENCE   78 AA;  8814 MW;  D609FADACD0B2BCC CRC64;

  Query Match             80.5%;  Score 33;  DB 1;  Length 78;
  Best Local Similarity   66.7%;  Pred. No. 19;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ::|||||||:
Db         12 VVIGVLVGL 20


RESULT 13
PTHB_ERWAM
ID   PTHB_ERWAM     STANDARD;      PRT;    333 AA.
AC   O32522;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   PTS system, glucitol/sorbitol-specific IIBC component (EIIBC-GUT)
DE   (Glucitol/sorbitol-permease IIBC component) (Phosphotransferase enzyme
DE   II, BC component) (EC 2.7.1.69) (EII-GUT).
GN   SRLE.
OS   Erwinia amylovora.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Erwinia.
OX   NCBI_TaxID=552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA7/74;
RX   MEDLINE=98098075; PubMed=9435786;
RA   Aldridge P., Metzger M., Geider K.;
RT   "Genetics of sorbitol metabolism in Erwinia amylovora and its
RT   influence on bacterial virulence.";
RL   Mol. Gen. Genet. 256:611-619(1997).
CC   -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent
CC       sugar phosphotransferase system (PTS), a major carbohydrate active
CC       -transport system. The IICD domains contain the sugar binding site
CC       and the transmembrane channel; the IIA domain contains the primary
CC       phosphorylation site (the donor is phospho-HPr); IIA transfers its
CC       phosphoryl group to the IIB domain which finally transfers it to
CC       the sugar.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: Contains 1 PTS EIIB domain.
CC   -!- SIMILARITY: Contains 1 PTS EIIC domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y14603; CAA74942.1; -.
DR   InterPro; IPR004702; Sorb_phosph_enII.
DR   Pfam; PF03612; EIIBC-GUT; 1.
DR   TIGRFAMs; TIGR00825; EIIBC-GUT; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Phosphorylation; Transmembrane; Inner membrane.
FT   DOMAIN        1      ?       EIIB.
FT   DOMAIN        ?    333       EIIC.
FT   TRANSMEM    160    180       POTENTIAL.
FT   TRANSMEM    191    211       POTENTIAL.
FT   TRANSMEM    220    240       POTENTIAL.
FT   TRANSMEM    243    263       POTENTIAL.
FT   TRANSMEM    271    291       POTENTIAL.
FT   TRANSMEM    304    324       POTENTIAL.
SQ   SEQUENCE   333 AA;  34292 MW;  6181206F04A61CAB CRC64;

  Query Match             80.5%;  Score 33;  DB 1;  Length 333;
  Best Local Similarity   87.5%;  Pred. No. 62;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 MIGVLVGV 9
              :|||||||
Db        253 VIGVLVGV 260


RESULT 14
Y390_MYCGE
ID   Y390_MYCGE     STANDARD;      PRT;    660 AA.
AC   Q49430; Q49332; Q49356;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical ATP-binding protein MG390.
```

```
GN   MG390.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 392-581 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- SIMILARITY: LIMITED, TO ABC TRANSPORTERS ATP-BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39721; AAC71618.1; -.
DR   EMBL; U02248; AAA03404.1; -.
DR   EMBL; U02218; AAA03372.1; -.
DR   PIR; B64243; B64243.
DR   TIGR; MG390; -.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR005074; Peptidase_C39.
DR   Pfam; PF00005; ABC_tran; 1.
DR   Pfam; PF03412; Peptidase_C39; 1.
KW   Hypothetical protein; ATP-binding; Complete proteome.
FT   NP_BIND     494    501       ATP (POTENTIAL).
FT   CONFLICT    392    392       L -> F (IN REF. 2).
SQ   SEQUENCE   660 AA;  76379 MW;  3CB9AEF9FFB89771 CRC64;

  Query Match             80.5%;  Score 33;  DB 1;  Length 660;
  Best Local Similarity   55.6%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ::||||:|:
Db        277 LIIGVLIGI 285


RESULT 15
ST42_ARATH
ID   ST42_ARATH     STANDARD;      PRT;   677 AA.
AC   Q8GYH8; Q9LHF7;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable sulfate transporter 4.2.
GN   SULTR4;2 OR AT3G12520 OR MQC3.34 OR T2E22.36.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Takahashi H., Watanabe-Takahashi A., Saito K., Yamaya T.;
RT   "cDNA for sulfate transporter Sultr4;2.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT   and BAC clones.";
RL   DNA Res. 7:217-221(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA   Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA   Hayashizaki Y., Shinozaki K.;
RT   "Arabidopsis thaliana full-length cDNA.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: H(+)/sulfate cotransporter that may play a role in the
CC       regulation of sulfate assimilation (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the SLC26A/SulP transporter (TC 2.A.53)
CC       family.
CC   -!- SIMILARITY: Contains 1 STAS domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB052775; BAB19761.1; -.
DR   EMBL; AP002047; BAB03159.1; -.
DR   EMBL; AC069474; AAG51021.1; -.
DR   EMBL; AK117615; BAC42271.1; -.
DR   InterPro; IPR002645; STAS.
```

```
DR   InterPro; IPR001902; Sulph_transpt.
DR   Pfam; PF01740; STAS; 1.
DR   Pfam; PF00916; Sulfate_transp; 1.
DR   TIGRFAMs; TIGR00815; sulP; 1.
DR   PROSITE; PS01130; SLC26A; 1.
DR   PROSITE; PS50801; STAS; 1.
KW   Transport; Symport; Sulfate transport; Transmembrane;
KW   Multigene family.
FT   DOMAIN        1     83       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     84    104       POTENTIAL.
FT   DOMAIN      105    108       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    109    129       POTENTIAL.
FT   DOMAIN      130    133       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    134    154       POTENTIAL.
FT   DOMAIN      155    161       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    162    182       POTENTIAL.
FT   DOMAIN      183    189       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    190    210       POTENTIAL.
FT   DOMAIN      211    241       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    242    262       POTENTIAL.
FT   DOMAIN      263    269       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    270    290       POTENTIAL.
FT   DOMAIN      291    318       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    319    339       POTENTIAL.
FT   DOMAIN      340    355       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    356    376       POTENTIAL.
FT   DOMAIN      377    392       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    393    413       POTENTIAL.
FT   DOMAIN      414    420       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    421    441       POTENTIAL.
FT   DOMAIN      442    459       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    460    480       POTENTIAL.
FT   DOMAIN      481    677       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      505    629       STAS.
FT   CONFLICT      7     22       MISSING (IN REF. 4).
SQ   SEQUENCE   677 AA;  74661 MW;  11C87626A781DB71 CRC64;

  Query Match             80.5%;  Score 33;  DB 1;  Length 677;
  Best Local Similarity   77.8%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              | ||||:||
Db        464 IEIGVLIGV 472


Search completed: August  6, 2004, 08:33:09
Job time : 15 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 6, 2004, 08:32:42 ; Search time 35 Seconds
(without alignments)
81.133 Million cell updates/sec

Title: US-09-458-302B-193
Perfect score: 41
Sequence: 1 IMIGVLVGV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	702	4	Q8N4D0	Q8n4d0 homo sapien
2	38	92.7	75	17	Q9YF39	Q9yf39 aeropyrum p
3	37	90.2	121	8	Q9B9D6	Q9b9d6 cilix glauc
4	37	90.2	126	8	Q7YCU2	Q7ycu2 synanthedon
5	37	90.2	447	16	Q8ZQJ4	Q8zqj4 salmonella
6	37	90.2	447	16	Q8Z839	Q8z839 salmonella
7	37	90.2	447	16	Q83T12	Q83t12 salmonella
8	36	87.8	125	8	Q7Y6W9	Q7y6w9 synanthedon
9	36	87.8	125	8	Q7Y6W8	Q7y6w8 synanthedon
10	36	87.8	346	13	Q7ZZ32	Q7zz32 brachydanio
11	36	87.8	378	11	Q9D400	Q9d400 mus musculu
12	36	87.8	539	16	Q8NNI8	Q8nni8 corynebacte
13	36	87.8	553	10	Q9LYQ0	Q9lyq0 arabidopsis
14	36	87.8	565	11	Q8C0X2	Q8c0x2 mus musculu
15	36	87.8	863	5	Q95VF8	Q95vf8 dictyosteli
16	36	87.8	893	5	Q9Y1Y3	Q9y1y3 ephydatia f
17	35	85.4	95	16	Q8UBX1	Q8ubx1 agrobacteri
18	35	85.4	104	8	Q34821	Q34821 ithomia sp.
19	35	85.4	104	8	Q34706	Q34706 heliconius
20	35	85.4	107	8	Q37413	Q37413 agraulis va
21	35	85.4	119	8	Q9B9D8	Q9b9d8 spodoptera
22	35	85.4	125	8	Q85GJ6	Q85gj6 cyclophora
23	35	85.4	192	17	O26661	O26661 methanobact
24	35	85.4	232	8	Q94UT1	Q94ut1 araschnia l
25	35	85.4	315	8	Q7YHL4	Q7yhl4 lepidopsoci
26	35	85.4	373	16	Q8FUF3	Q8fuf3 corynebacte
27	35	85.4	402	16	Q8RFT3	Q8rft3 fusobacteri
28	35	85.4	423	5	Q20396	Q20396 caenorhabdi
29	35	85.4	428	16	Q9I2U4	Q9i2u4 pseudomonas
30	35	85.4	563	16	Q8NLT6	Q8nlt6 corynebacte
31	35	85.4	614	16	Q7UUY6	Q7uuy6 rhodopirell
32	34	82.9	58	16	Q8YN92	Q8yn92 anabaena sp
33	34	82.9	94	8	Q37613	Q37613 phyciodes s
34	34	82.9	103	8	Q37456	Q37456 danaus plex
35	34	82.9	104	8	Q37692	Q37692 vanessa ata
36	34	82.9	104	8	Q37418	Q37418 battus phil
37	34	82.9	104	8	Q37537	Q37537 libytheana
38	34	82.9	104	8	Q37367	Q37367 anaea andri
39	34	82.9	104	8	Q37555	Q37555 macrosoma s
40	34	82.9	104	8	Q37609	Q37609 pontia prot
41	34	82.9	104	8	Q37467	Q37467 enodia port
42	34	82.9	104	8	Q37670	Q37670 speyeria at
43	34	82.9	120	8	Q9B9D9	Q9b9d9 agrotis seg
44	34	82.9	120	8	Q9B9D0	Q9b9d0 idaea stram
45	34	82.9	125	8	Q9B9C9	Q9b9c9 idaea avers

ALIGNMENTS

RESULT 1
Q8N4D0
ID Q8N4D0 PRELIMINARY; PRT; 702 AA.
AC Q8N4D0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Kidney;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC034671; AAH34671.1; -.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; ig; 6.
DR SMART; SM00409; IG; 7.
DR SMART; SM00408; IGc2; 6.
DR PROSITE; PS00019; ACTININ_1; 3.
DR PROSITE; PS50835; IG_LIKE; 6.
KW Immunoglobulin domain.
SQ SEQUENCE 702 AA; 76781 MW; 97CCFB7399A0B05A CRC64;

Query Match 100.0%; Score 41; DB 4; Length 702;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IMIGVLVGV 9
|||||||||
Db 691 IMIGVLVGV 699

```
RESULT 2
Q9YF39
ID   Q9YF39      PRELIMINARY;      PRT;    75 AA.
AC   Q9YF39;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein APES012.
GN   APES012.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC   Desulfurococcaceae; Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000059; BAA79357.1; -.
DR   PIR; A72733; A72733.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   75 AA;  8155 MW;  C3601D24F88A68D9 CRC64;

  Query Match             92.7%;  Score 38;  DB 17;  Length 75;
  Best Local Similarity   66.7%;  Pred. No. 17;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ||:|||:|:
Db         30 IMVGVLIGI 38


RESULT 3
Q9B9D6
ID   Q9B9D6      PRELIMINARY;      PRT;   121 AA.
AC   Q9B9D6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   NADH dehydrogenase subunit 1 (EC 1.6.5.3) (NADH-ubiquinone
DE   oxidoreductase chain 1) (Fragment).
GN   ND1.
OS   Cilix glaucata.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Drepanoidea;
OC   Drepanidae; Drepaninae; Cilix.
OX   NCBI_TaxID=104441;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Abraham D., Ryrholm N., Wittzell H., Scoble M.J., Holloway J.D.,
RA   Lofstedt C.;
RT   "Molecular phylogeny of the subfamilies in Geometridae (Geometroidea:
RT   Lepidoptera).";
RL   Mol. Phylogenet. Evol. 0:0-0(2001).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
DR   EMBL; AF178859; AAK00986.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005739; C:mitochondrion; IEA.
DR   GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001694; Resp_NADH_dh1.
DR   Pfam; PF00146; NADHdh; 1.
DR   PROSITE; PS00667; COMPLEX1_ND1_1; 1.
KW   NAD; Oxidoreductase; Transmembrane; Ubiquinone; Mitochondrion.
FT   NON_TER       1      1
FT   NON_TER     121    121
SQ   SEQUENCE   121 AA;  14016 MW;  DC96C584C246D73B CRC64;

  Query Match             90.2%;  Score 37;  DB 8;  Length 121;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|||:||
Db         14 LMVGVLIGV 22


RESULT 4
Q7YCU2
ID   Q7YCU2      PRELIMINARY;      PRT;   126 AA.
AC   Q7YCU2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NADH dehydrogenase subunit 1 (Fragment).
OS   Synanthedon spheciformis (white-barred clearwing).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sesioidea;
OC   Sesiidae; Sesiinae; Synanthedon.
OX   NCBI_TaxID=233843;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK5;
RA   Kallies A.;
RT   "Phylogeny of sesiid Taxa.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY304167; AAP84246.1; -.
KW   Mitochondrion.
FT   NON_TER     126    126
SQ   SEQUENCE   126 AA;  14304 MW;  E606AB4CB0BCA50F CRC64;

  Query Match             90.2%;  Score 37;  DB 8;  Length 126;
  Best Local Similarity   77.8%;  Pred. No. 42;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:||||||
Db         16 LMLGVLVGV 24


RESULT 5
Q8ZQJ4
ID   Q8ZQJ4      PRELIMINARY;      PRT;   447 AA.
AC   Q8ZQJ4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative inner membrane protein, homology to SgaT from Vibrio.
GN   STM0884.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
```

```
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008737; AAL19819.1; -.
DR   InterPro; IPR007333; SgaT_UlaA.
DR   Pfam; PF04215; SgaT_UlaA; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   447 AA;  48656 MW;  886B4E3FA2601154 CRC64;

  Query Match             90.2%;  Score 37;  DB 16;  Length 447;
  Best Local Similarity   77.8%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |:||:||||
Db         24 ILIGILVGV 32


RESULT 6
Q8Z839
ID   Q8Z839      PRELIMINARY;      PRT;   447 AA.
AC   Q8Z839;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Possible transport protein.
GN   STY0917.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627268; CAD05323.1; -.
DR   InterPro; IPR007333; SgaT_UlaA.
DR   Pfam; PF04215; SgaT_UlaA; 1.
KW   Complete proteome.
SQ   SEQUENCE   447 AA;  48638 MW;  AECE4B7D47640976 CRC64;

  Query Match             90.2%;  Score 37;  DB 16;  Length 447;
  Best Local Similarity   77.8%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |:||:||||
Db         24 ILIGILVGV 32


RESULT 7
Q83T12
ID   Q83T12      PRELIMINARY;      PRT;   447 AA.
AC   Q83T12;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Possible transport protein.
GN   T2012.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
DR   EMBL; AE016840; AAO69624.1; -.
DR   InterPro; IPR007333; SgaT_UlaA.
DR   Pfam; PF04215; SgaT_UlaA; 1.
SQ   SEQUENCE   447 AA;  48656 MW;  4BDA1A3D1362A3B0 CRC64;

  Query Match             90.2%;  Score 37;  DB 16;  Length 447;
  Best Local Similarity   77.8%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |:||:||||
Db         24 ILIGILVGV 32


RESULT 8
Q7Y6W9
ID   Q7Y6W9      PRELIMINARY;      PRT;   125 AA.
AC   Q7Y6W9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NADH dehydrogenase subunit 1 (Fragment).
OS   Synanthedon culiciformis (large red-belted clearwing).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sesioidea;
OC   Sesiidae; Sesiinae; Synanthedon.
OX   NCBI_TaxID=233842;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK3, AK11, and AK4;
RA   Kallies A.;
RT   "Phylogeny of sesiid Taxa.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY304162; AAP84241.1; -.
DR   EMBL; AY304163; AAP84242.1; -.
DR   EMBL; AY304166; AAP84245.1; -.
KW   Mitochondrion.
FT   NON_TER     125    125
SQ   SEQUENCE   125 AA;  14294 MW;  960B26BE1C96656E CRC64;

  Query Match             87.8%;  Score 36;  DB 8;  Length 125;
  Best Local Similarity   66.7%;  Pred. No. 63;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|||:||
Db         16 LMLGVLIGV 24


RESULT 9
Q7Y6W8
ID   Q7Y6W8      PRELIMINARY;      PRT;   125 AA.
AC   Q7Y6W8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NADH dehydrogenase subunit 1 (Fragment).
OS   Synanthedon pamphyla.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sesioidea;
OC   Sesiidae; Sesiinae; Synanthedon.
OX   NCBI_TaxID=233844;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AK1, and AK8;
RA   Kallies A.;
RT   "Phylogeny of sesiid Taxa.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY304164; AAP84243.1; -.
DR   EMBL; AY304165; AAP84244.1; -.
KW   Mitochondrion.
FT   NON_TER     125    125
SQ   SEQUENCE   125 AA;  14294 MW;  960B26BE1C96656E CRC64;

  Query Match          87.8%;  Score 36;  DB 8;  Length 125;
  Best Local Similarity   66.7%;  Pred. No. 63;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|||:||
Db         16 LMLGVLIGV 24


RESULT 10
Q7ZZ32
ID   Q7ZZ32      PRELIMINARY;      PRT;   346 AA.
AC   Q7ZZ32;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SI:zC12P8.3 (Novel protein similar to human 5-hydroxytryptamine
DE   (Serotonin) receptor 5A (HTR5A)).
GN   SI:ZC12P8.3.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Corby N.;
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL772146; CAD61100.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
SQ   SEQUENCE   346 AA;  39412 MW;  B554D1BC1E74413E CRC64;

  Query Match          87.8%;  Score 36;  DB 13;  Length 346;
  Best Local Similarity   55.6%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|:|:|:||
Db        274 LMVGILIGV 282


RESULT 11
Q9D400
ID   Q9D400      PRELIMINARY;      PRT;   378 AA.
AC   Q9D400;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   4933425K02Rik protein.
GN   4933425K02RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK016917; BAB30495.1; -.
DR   MGD; MGI:1921696; 4933425K02Rik.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0015299; F:solute:hydrogen antiporter activity; IEA.
DR   GO; GO:0006885; P:regulation of pH; IEA.
DR   InterPro; IPR006153; Na_H_porter.
DR   Pfam; PF00999; Na_H_Exchanger; 1.
SQ   SEQUENCE   378 AA;  41808 MW;  2F3AA347A2DE6188 CRC64;

  Query Match          87.8%;  Score 36;  DB 11;  Length 378;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ::||||||:
Db        341 VLIGVLVGI 349


RESULT 12
Q8NNI8
ID   Q8NNI8      PRELIMINARY;      PRT;   539 AA.
AC   Q8NNI8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Predicted permease.
GN   CGL2211.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005281; BAB99604.1; -.
DR   GO; GO:0008324; F:cation transporter activity; IEA.
DR   GO; GO:0006813; P:potassium ion transport; IEA.
DR   InterPro; IPR006037; TrkAC.
DR   InterPro; IPR006512; YidE_YbjL.
DR   Pfam; PF02080; TrkA-C; 1.
DR   TIGRFAMs; TIGR01625; YidE_YbjL_dupl; 2.
KW   Complete proteome.
SQ   SEQUENCE   539 AA;  57150 MW;  EE6E907F6D29FD7B CRC64;

  Query Match          87.8%;  Score 36;  DB 16;  Length 539;
  Best Local Similarity   77.8%;  Pred. No. 2.4e+02;
```

```
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              :|||||||:
Db        369 LMIGVLVGM 377


RESULT 13
Q9LYQ0
ID   Q9LYQ0      PRELIMINARY;      PRT;   553 AA.
AC   Q9LYQ0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   T28J14_90.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL163652; CAB87271.1; -.
DR   PIR; T48486; T48486.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004672; F:protein kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007090; LRR_plant.
DR   InterPro; IPR000719; Prot_kinase.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein; ATP-binding; Transferase.
SQ   SEQUENCE   553 AA;  61666 MW;  83149DBFE099D39D CRC64;

  Query Match             87.8%;  Score 36;  DB 10;  Length 553;
  Best Local Similarity   77.8%;  Pred. No. 2.5e+02;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              |::||||||
Db        232 IIVGVLVGV 240


RESULT 14
Q8C0X2
ID   Q8C0X2      PRELIMINARY;      PRT;   565 AA.
AC   Q8C0X2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical glutamic acid-rich region/Na+/H+ exchanger containing
DE   protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK029525; BAC26494.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0015299; F:solute:hydrogen antiporter activity; IEA.
DR   GO; GO:0006885; P:regulation of pH; IEA.
DR   InterPro; IPR006153; Na_H_porter.
DR   Pfam; PF00999; Na_H_Exchanger; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   565 AA;  61957 MW;  7ECBC2E03DC90655 CRC64;

  Query Match             87.8%;  Score 36;  DB 11;  Length 565;
  Best Local Similarity   66.7%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IMIGVLVGV 9
              ::||||||:
Db        341 VLIGVLVGI 349


RESULT 15
Q95VF8
ID   Q95VF8      PRELIMINARY;      PRT;   863 AA.
AC   Q95VF8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Chloride channel protein ClcA.
GN   CLCA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KAX3;
RA   Wang C.W., Liu C.I., Chang W.T.;
RT   "Molecular analyses and functional studies of chloride channel protein
RT   ClcA in Dictyostelium.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF414428; AAL07438.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005247; F:voltage-gated chloride channel activity; IEA.
DR   GO; GO:0006821; P:chloride transport; IEA.
DR   InterPro; IPR000644; CBS_domain.
DR   InterPro; IPR001807; Cl-channel_volt.
DR   InterPro; IPR006311; Tat.
DR   Pfam; PF00571; CBS; 2.
DR   Pfam; PF00654; voltage_CLC; 1.
DR   PRINTS; PR00762; CLCHANNEL.
DR   SMART; SM00116; CBS; 1.
DR   TIGRFAMs; TIGR01409; TAT_signal_seq; 1.
SQ   SEQUENCE   863 AA;  97298 MW;  575CEE036AE0A435 CRC64;

  Query Match             87.8%;  Score 36;  DB 5;  Length 863;
  Best Local Similarity   87.5%;  Pred. No. 3.7e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 MIGVLVGV 9
              |||||||:
Db        133 MIGVLVGI 140


Search completed: August  6, 2004, 08:35:00
Job time : 37 secs
```